

Breves de Algebra Lineal

Salvador Pintos y Guido Urdaneta

Revisado por C. Bustos, S. Luzardo, G.Meza, M. Quintero, A. Riera y J. Urdaneta

Índice general

Capítulo 1. SISTEMAS DE ECUACIONES	5
1.1. Geometría de los Sistemas Lineales	5
1.2. Eliminación Gaussiana	6
Capítulo 2. MATRICES	7
2.1. Suma de Matrices	7
2.2. Producto de un escalar y una matriz	7
2.3. Producto Matricial	7
2.4. Notación matricial de sistemas de ecuaciones lineales	8
2.5. Matriz Identidad	9
2.6. Inversa	9
2.7. Traspuesta	10
2.8. Matrices Elementales	10
2.9. Matrices de Permutación	12
Capítulo 3. ESPACIOS VECTORIALES	13
3.1. Independencia y dependencia lineal	13
3.2. Generadores, Bases y Dimensión	14
3.3. Subespacios vectoriales	14
Capítulo 4. TRANSFORMACIONES LINEALES	16
Capítulo 5. ESPACIOS FUNDAMENTALES	18
5.1. Espacio de las filas de A , $R(A^T)$	18
5.2. Espacio Nulo de A , $N(A)$	19
5.3. Espacio de las columnas de A , $R(A)$	20
5.4. Espacio Nulo a la izquierda de A , $N(A^T)$	20
5.5. Bases complementarias	21
5.6. Solución de sistemas	21
Capítulo 6. ORTOGONALIDAD	24
6.1. Producto escalar	24
6.2. Algunas Propiedades	24
6.3. Vectores ortogonales	24
6.4. Espacios fundamentales	25
6.5. Ángulos entre vectores	25
6.6. Proyecciones sobre líneas	26
6.7. Proyecciones sobre espacios y aproximaciones con mínimos cuadrados	26
6.8. Bases ortogonales	28
Capítulo 7. DETERMINANTES	31
7.1. Propiedades	31
7.2. El determinante como medida de volumen	32

7.3. Regla de Cramer	33
Capítulo 8. VALORES Y VECTORES PROPIOS	34
8.1. Propiedades	34
8.2. Forma Diagonal	35
8.3. Aplicaciones	36
8.4. Matrices simétricas	38
8.5. Formas cuadráticas	39
Capítulo 9. EXTREMOS DE FUNCIONES DE VARIAS VARIABLES	42
9.1. Teorema de Taylor	42
Capítulo 10. DESCOMPOSICIÓN EN VALORES SINGULARES (SVD)	46
10.1. Introducción	46
10.2. Aplicación a compresión de datos	47
10.3. Construcción de la SVD	47

CAPÍTULO 1

SISTEMAS DE ECUACIONES

El problema central del álgebra lineal es la solución de sistemas de ecuaciones lineales. Una ecuación lineal tiene la forma $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$ donde los x_i son incógnitas y b y los a_i son escalares constantes. Una solución de esta ecuación es un conjunto de valores de las incógnitas x_i que satisface la igualdad.

Un sistema de ecuaciones lineales es un conjunto de ecuaciones lineales $\{e_1, e_2, \dots, e_m\}$ y su solución es la intersección de las soluciones individuales de cada una de las ecuaciones. Si la intersección es vacía entonces el sistema no tiene solución.

Dos sistemas de ecuaciones son equivalentes si tienen el mismo conjunto solución.

Cuando un sistema tiene solución se dice que es *compatible*. Si la solución es única entonces es *determinado*. Si el número de soluciones es infinito entonces es indeterminado.

Cuando un sistema no tiene solución se dice que es *incompatible*.

El caso más importante y simple ocurre cuando el número de incógnitas es igual al número de ecuaciones. Existen dos métodos bien definidos para resolver este problema. El primero es la *eliminación gaussiana*, en la cual múltiplos de la primera ecuación se restan de las otras ecuaciones, lo cual deja un sistema más pequeño con $n - 1$ ecuaciones y $n - 1$ incógnitas. El proceso se repite hasta que quede una sola ecuación y una sola incógnita, la cual puede ser resuelta de inmediato; luego resulta fácil ir hacia atrás y encontrar las otras incógnitas en orden inverso. El segundo método introduce el concepto de *determinante*. Existe una fórmula exacta llamada regla de Cramer, la cual permite resolver cada incógnita como el cociente de dos determinantes.

En la práctica, la regla de Cramer es un desastre y la eliminación es el algoritmo utilizado constantemente para resolver sistemas grandes de ecuaciones.

1.1. Geometría de los Sistemas Lineales

Considérese el sistema de ecuaciones:

$$\begin{aligned}2x + 3y &= 8 \\4x - 5y &= -6\end{aligned}$$

1.1.1. Visión Geométrica 1. La primera visión geométrica está basada en las filas. Cada ecuación representa una recta. Resolver el sistema consiste en encontrar la intersección de todas las rectas.

1.1.2. Visión Geométrica 2. El segundo enfoque se concentra en las columnas del sistema. Las dos ecuaciones son realmente una ecuación vectorial:

$$x \begin{bmatrix} 2 \\ 4 \end{bmatrix} + y \begin{bmatrix} 3 \\ -5 \end{bmatrix} = \begin{bmatrix} 8 \\ -6 \end{bmatrix}$$

Resolver el sistema consiste en encontrar la combinación lineal de los vectores columna al lado izquierdo que produce el vector al lado derecho.

1.2. Eliminación Gaussiana

La eliminación gaussiana pretende, partiendo de un sistema, llegar a un sistema equivalente donde encontrar la solución utilizando un esquema de sustitución hacia atrás sea bien sencillo.

La eliminación gaussiana se basa en en los siguientes teoremas::

TEOREMA. *Teorema Fundamental de Sistemas.*

$$S = \begin{cases} e_1 \\ e_2 \\ \vdots \\ e_k \\ \vdots \\ e_m \end{cases} \text{ es equivalente a } S' = \begin{cases} e_1 \\ e_2 \\ \vdots \\ e'_k = \sum_{j=1}^m \alpha_j e_j \\ \vdots \\ e_m \end{cases} \quad \text{si } \alpha_k \neq 0.$$

La k -ésima fila del sistema fue sustituida por una combinación lineal de las demás filas.

TEOREMA. *Un sistema S y el sistema S' que resulta de intercambiar dos filas cualesquiera de S son equivalentes.*

1.2.1. Pivotes. Se denomina pivote al primer elemento distinto de cero en una fila después de realizar el proceso de escalerización. Los pivotes tienen las siguientes características:

1. Debajo de un pivote hay ceros.
2. Todo pivote de una fila está a la derecha del pivote de la fila anterior.

CAPÍTULO 2

MATRICES

Una matriz es un arreglo rectangular de números¹. Las líneas horizontales de una matriz se denominan *filas* y las verticales *columnas*. Una matriz con m filas y n columnas es referida como una matriz de $m \times n$ (m por n). Por ejemplo, la siguiente es una matriz de 4×3 :

$$\begin{bmatrix} 2 & 2 & 5 \\ -1 & 1 & -2 \\ 3 & 2 & -3 \\ 5 & -8 & 0 \end{bmatrix}$$

El elemento de una matriz A ubicado en la fila i y la columna j se suele denotar A_{ij} o a_{ij} .

2.1. Suma de Matrices

La suma de dos matrices A y B está definida sólo si ambas tienen el mismo número de filas y columnas.

DEFINICIÓN. Si A y B son matrices de $m \times n$, entonces su suma $A + B$ es una matriz $m \times n$ dada por

$$(A + B)_{ij} = a_{ij} + b_{ij}$$

para cada par i y j .

2.2. Producto de un escalar y una matriz

DEFINICIÓN. Si c es un escalar y A es una matriz de $m \times n$ entonces el producto cA es una matriz de $m \times n$ dada por

$$(cA)_{ij} = ca_{ij}$$

para todo par i, j .

2.3. Producto Matricial

El producto de dos matrices A y B está definido si el número de columnas de A es igual al número de filas de B .

DEFINICIÓN. Si A es una matriz de $m \times p$ y B es una matriz de $p \times n$, entonces su producto AB es una matriz de $m \times n$ dada por

$$(AB)_{ij} = \sum_{k=1}^p a_{ik}b_{kj}$$

para cada par i y j .

¹En nuestro caso, a menos que se indique lo contrario, se tratará de números reales. Más generalmente, los elementos de una matriz pueden ser miembros de un anillo.

Esta noción de multiplicación es importante porque define la multiplicación por una matriz como una transformación lineal y el producto de dos matrices como la composición de dos transformaciones lineales.

La multiplicación de matrices es:

- Asociativa: $(AB)C = A(BC)$
- Distributiva: $A(B + C) = AB + AC$, $(A + B)C = AC + BC$

En general, la multiplicación de matrices no es conmutativa: $AB \neq BA$

El producto AB de dos matrices puede verse de varias maneras:

1. Cada elemento de AB es el producto de una fila y una columna:

$$(AB)_{ij} = A_{i.}B_{.j}$$

2. Cada columna de AB es el producto de una matriz y una columna:

$$(AB)_{.j} = AB_{.j}$$

Nótese que el producto de A y una columna de B puede interpretarse como una combinación lineal de las columnas de A utilizando como coeficientes las coordenadas del vector columna de B considerado.

3. Cada fila de AB es el producto de una fila y una matriz:

$$(AB)_{i.} = A_{i.}B$$

Nótese que el producto una fila de A y la matriz B puede interpretarse como una combinación lineal de las filas de B utilizando como coeficientes las coordenadas del vector fila de A considerado.

4. AB puede verse como la suma de las matrices resultantes de multiplicar columnas y filas:

$$AB = \sum_{k=1}^p A_{.k}B_{k.}$$

donde p es el número de columnas de A y el número de filas de B .

2.4. Notación matricial de sistemas de ecuaciones lineales

Considere el siguiente sistema de ecuaciones:

$$\begin{aligned} 2x_1 + x_2 + x_3 &= 4 \\ 4x_1 - 6x_2 + x_3 &= 15 \\ 3x_1 - x_2 - x_3 &= 6 \end{aligned}$$

Este sistema de ecuaciones se puede representar como $Ax = b$, siendo $A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 1 \\ 3 & -1 & -1 \end{bmatrix}$,

$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ y $b = \begin{bmatrix} 4 \\ 15 \\ 6 \end{bmatrix}$. Nótese que x y b son vectores columnas, pero al mismo tiempo podemos interpretarlos como matrices de una sola columna.

La solución de este sistema es el vector $x = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$

Esta notación nos permite interpretar el sistema de ecuaciones de dos maneras:

1. Por filas: cada fila de Ax es combinación lineal de las variables.
2. Por columnas: El producto Ax es una combinación lineal de las columnas de A .

Al aplicar eliminación gaussiana, el sistema $Ax = b$ se transforma en uno equivalente de la forma $Ux = c$ donde U es una matriz triangular superior. El número de pivotes en U se conoce como el *rango* de la matriz A .

2.5. Matriz Identidad

DEFINICIÓN. La matriz cuadrada $I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$ con unos en la diagonal y ceros en las otras posiciones se conoce como matriz identidad y tiene la característica de que $A_{m \times n} I_n = A$.

2.6. Inversa

DEFINICIÓN. La inversa de una matriz cuadrada A es una matriz A^{-1} , tal que $A^{-1}A = I$ y $AA^{-1} = I$.

Si se conoce la inversa de una matriz A entonces para resolver el sistema $Ax = b$ basta multiplicar por A^{-1} y nos queda $A^{-1}Ax = x = A^{-1}b$.

Algunas Propiedades:

Cuando una matriz es invertible la inversa debe ser única. Si existieran dos matrices diferentes B y C tales que $BA = I$ y $AC = I$ entonces nos quedaría que $B(AC) = BI = B$, pero también $B(AC) = (BA)C = C$, es decir $B = B(AC) = C$. Por lo tanto B y C no pueden ser diferentes.

$(A^{-1})^{-1} = A$, cumple con la definición.

$(AB)^{-1} = B^{-1}A^{-1}$, se verifica multiplicando ambos lados por AB .

$$\begin{aligned} (B^{-1}A^{-1})(AB) &= B^{-1}A^{-1}AB = B^{-1}IB = B^{-1}B = I \\ (AB)(B^{-1}A^{-1}) &= ABB^{-1}A^{-1} = AIA^{-1} = AA^{-1} = I \end{aligned}$$

Similarmente, $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$.

Una matriz cuadrada A de $n \times n$ es invertible si y sólo si su forma escalerizada tiene n pivotes. Si la forma escalerizada tiene algún cero en la diagonal entonces no es invertible y se dice que la matriz A es singular.

Para calcular la inversa, se procede de manera similar a la utilizada para resolver un sistema $Ax = b$, pero en lugar de utilizar un único b , se utilizan los vectores columnas de la matriz identidad. La eliminación se debe extender hasta que la matriz a la izquierda quede igual a la identidad (eliminación Gauss-Jordan), entonces la matriz a la derecha será la inversa A^{-1} .

2.7. Traspuesta

DEFINICIÓN. La traspuesta de una matriz A se denota A^T y sus filas corresponden a las columnas de A y sus columnas a las filas de A .

$$(A^T)_{ij} = A_{ji}$$

Propiedades:

$(AB)^T = B^T A^T$, se verifica al observar que $((AB)^T)_{ij}$ es $(AB)_{ji}$ el cual es el producto de la fila j de A y la columna i de B , mientras que $(B^T A^T)_{ij}$ es el producto de la fila i de B^T (la cual es la columna i de B) y la columna j de A^T (que es la fila j de A).

$(A + B)^T = A^T + B^T$, trivial.

$(A^{-1})^T = (A^T)^{-1}$, se verifica al hacer $AA^{-1} = I$ y $A^{-1}A = I$, trasponiendo, por un lado $I^T = I$, $(AA^{-1})^T = (A^{-1})^T A^T = I$ y $(A^{-1}A)^T = A^T (A^{-1})^T = I$, lo que hace que $(A^{-1})^T$ sea la inversa de A^T .

Geoméricamente, si A representa una transformación de un espacio E^n a uno E^m , la traspuesta representa una transformación del espacio E^m al E^n (aunque no necesariamente una transformación inversa).

Una matriz en la que se cumple $A^T = A$ se denomina *matriz simétrica*. Estas matrices necesariamente son cuadradas y la simetría ocurre con respecto a la diagonal. En una matriz simétrica cada elemento $a_{ij} = a_{ji}$.

2.8. Matrices Elementales

DEFINICIÓN. Una matriz subdiagonal tiene unos en la diagonal principal, un valor distinto de cero en una posición (i, j) con $i > j$ y ceros en las demás posiciones. Multiplicar una matriz elemental E con el valor l en la posición (i, j) por una matriz A produce una matriz que resulta de sumar un múltiplo l de la ecuación (fila) j a la ecuación i .

EJEMPLO. Sea $A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix}$, para poner un cero en la posición $(2, 1)$ podemos restar dos

veces la fila 1 de la fila 2. Esto es equivalente a multiplicar la matriz elemental $E = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

$$\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ -2 & 7 & 2 \end{bmatrix}$$

Para poner cero en la posición $(3, 1)$ sumamos la fila 1 a la 3, lo cual equivale a multiplicar la matriz

elemental $F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ -2 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 8 & 3 \end{bmatrix}$$

Finalmente, para tener la forma escalerizada se debe poner un cero en la posición (3, 2). Para esto sumamos la ecuación 2 a la ecuación 3, lo cual equivale a multiplicar por la matriz elemental

$$G = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 8 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

Nótese que todo el procedimiento de escalerización es la multiplicación sucesiva de matrices elementales:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

La multiplicación de matrices elementales produce siempre una matriz triangular inferior. En este caso:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

La inversa de una matriz elemental es muy fácil de calcular. Básicamente consiste en multiplicar por -1 el elemento no cero de la posición (i, j) ($i > j$). Es fácil verificar que

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

El proceso inverso a la escalerización sería multiplicar la matriz escalerizada sucesivamente por las inversas de las matrices elementales usadas para escalerizar.

Por ejemplo, si $GFEA = U$ (U triangular superior, G, F, E elementales), entonces $A = E^{-1}F^{-1}G^{-1}U = LU$, siendo L una matriz triangular inferior. En el ejemplo,

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

La ecuación $A = LU$ proporciona una manera conveniente de resolver un sistema.

Si $Ax = b$, entonces $LUx = b$. Si se hace $Ux = y$, entonces se puede resolver fácilmente $Ly = b$ y luego $Ux = y$. Ambos sistemas son fáciles de resolver porque L y U son triangulares.

En ocasiones, la factorización LU se escribe como LDU donde L y U tienen unos en la diagonal y D es una matriz diagonal con los pivotes.

2.9. Matrices de Permutación

DEFINICIÓN. Una matriz de permutación P es una matriz que tiene las mismas filas de la matriz identidad, pero no necesariamente en el mismo orden.

Una matriz de permutación resulta del intercambio de filas de una matriz identidad. El efecto de premultiplicar una matriz A por esta matriz es intercambiar las filas en A .

$$\text{Ejemplo: } P_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \text{ multiplicada por } A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 6 \\ 0 & 3 & 2 \end{bmatrix} \text{ produce } PA = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 6 \end{bmatrix}.$$

La realización de varios intercambios de filas puede efectuarse mediante la multiplicación sucesiva de matrices de permutación. Por ejemplo, para intercambiar la fila 1 con la fila 3 y luego la fila 2 con la fila 3 basta con multiplicar el resultado de las matrices de permutación correspondientes:

$$P_{23}P_{13} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = P$$

Nótese que una matriz de permutación P tiene las mismas filas que la identidad, en algún orden. De hecho, $P = I$ es la matriz de permutación más simple y común (no intercambia nada). El producto de dos matrices de permutación es otra matriz de permutación. El producto de matrices de permutación no conmuta. En el ejemplo anterior, $P_{13}P_{23}$ daría una matriz P diferente.

Siempre pueden hacerse los intercambios de filas antes de proceder a la eliminación, aunque normalmente no se sabe de antemano cuales filas hay que intercambiar. El hecho de que al aplicar el procedimiento de escalerización, normalmente los intercambios de filas se realizan en pasos intermedios, realmente no importa cuando éstos se realicen. Podemos asumir que se realizan desde un principio y trabajar con la matriz PA , que puede ser escalerizada mediante la multiplicación sucesiva de matrices elementales y por lo tanto puede ser factorizada como LU .

En resumen, en el caso no singular existe una matriz P que reordena las filas de A para evitar ceros en las posiciones de los pivotes. En este caso:

- $Ax = b$ tiene una solución única
- La solución se consigue mediante eliminación con intercambios de filas
- Con las filas reordenadas de antemano, PA puede ser factorizada como LU

En el caso singular (A no invertible) no hay manera de poner números distintos de cero en todas las posiciones de los pivotes.

CAPÍTULO 3

ESPACIOS VECTORIALES

DEFINICIÓN. Un espacio vectorial es un triplete $(V, +, \cdot)$. Los elementos del conjunto V se llaman vectores, $+$ es una operación de suma entre vectores $(+ : V \times V \longrightarrow V)$ y \cdot es una operación de multiplicación de un escalar por un vector que produce como resultado un vector; en nuestro caso, a menos que explícitamente se indique lo contrario, los escalares pertenecerán al conjunto de los números reales, entonces $\cdot : \mathbb{R} \times V \longrightarrow V$.

Un espacio vectorial debe cumplir con las siguientes propiedades:

1. $(u + v) + w = u + (v + w)$
2. $uv = vu$
3. Existe un único “vector cero” tal que $v + 0 = v$, $v \cdot 0 = 0$ para todo v .
4. Para todo vector v existe un único vector $-v$ tal que $v + (-v) = 0$
5. $1 \cdot v = v$
6. $(c_1 c_2)v = c_1(c_2 v)$
7. $c(x + y) = cx + cy$
8. $(c_1 + c_2)x = c_1 x + c_2 x$

Ejemplos de espacios vectoriales:

1. Vectores columna en \mathbb{R}^n .
2. Espacio de dimensión infinita \mathbb{R}^∞ .
3. Espacio de matrices de 3×2 , o de $m \times n$.
4. Espacio de polinomios de grado n .
5. Espacio de funciones continuas en el intervalo $[0, 1]$

Los espacios vectoriales siempre son infinitos (excepto el caso especial de los espacios vectoriales que sólo contienen un vector nulo).

3.1. Independencia y dependencia lineal

Una expresión de la forma $\sum_{i=1}^n c_i v_i$ donde los c_i son escalares y los v_i son vectores se denomina *combinación lineal*.

Un conjunto de vectores $\{v_1, v_2, \dots, v_n\}$ es *linealmente independiente* (LI) si se cumple que

$$\sum_{i=1}^n c_i v_i = 0 \rightarrow \forall i (c_i = 0)$$

En un sistema cuadrado homogéneo la solución es cero y por lo tanto los vectores son linealmente independientes.

Un conjunto de vectores es *linealmente dependiente* (LD) cuando al menos uno de los vectores es función del resto.

Por ejemplo: $\{1, x, x^2, 3x - 2x^2\}$ es LD, ya que un vector se puede escribir como combinación lineal de los otros.

Nótese que la independencia lineal es una propiedad del conjunto de vectores y no de los vectores individualmente.

3.2. Generadores, Bases y Dimensión

Un conjunto de vectores $\{w_1, w_2, \dots, w_n\}$ es un *generador* de un espacio vectorial si cualquier vector del espacio se puede expresar como una combinación lineal de ellos.

$$\forall v (\exists c_1, c_2, \dots, c_n (v = \sum_{i=1}^n c_i w_i))$$

Lo más deseable de un conjunto generador es que sea LI. En este caso el generador recibe el nombre de *base*.

Cuando se trata de una base, para cualquier vector v , la combinación lineal de vectores que produce v es única. Suponga que $\{v_1, v_2, \dots, v_n\}$ es una base y que el vector v se puede representar de dos maneras:

$$\begin{aligned} v &= a_1 v_1 + a_2 v_2 + \dots + a_n v_n \\ v &= b_1 v_1 + b_2 v_2 + \dots + b_n v_n \end{aligned}$$

si restamos ambas ecuaciones nos queda

$$0 = v_1(a_1 - b_1) + v_2(a_2 - b_2) + \dots + v_n(a_n - b_n)$$

como la base siempre es LI entonces $a_i - b_i = 0$ y $a_i = b_i$ para todo i . Por lo tanto, sólo existe una manera de escribir v como combinación lineal de los vectores de la base. Si $v = a_1 v_1 + a_2 v_2 + \dots + a_n v_n$ entonces los escalares a_i son las coordenadas de v con respecto a la base $\{v_1, v_2, \dots, v_n\}$.

El número de elementos de la base recibe el nombre de *dimensión del espacio vectorial*. Un espacio vectorial tiene infinitas bases, pero la dimensión siempre es la misma. La dimensión es una propiedad del espacio vectorial. Siempre que una matriz cuadrada sea invertible, sus columnas son linealmente independientes y constituyen una base para \mathbb{R}^n .

Por ejemplo, como la matriz $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ es invertible, entonces todo vector en \mathbb{R}^2 se puede representar como una combinación lineal de los vectores $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$ y $\begin{bmatrix} 2 \\ 4 \end{bmatrix}$.

3.3. Subespacios vectoriales

DEFINICIÓN. Un *subespacio* de un espacio vectorial es un subconjunto no vacío que satisface dos requerimientos:

1. Si se tienen dos vectores u y v en el subespacio, su suma $u + v$ también está en el subespacio.
2. Si se multiplica un vector v en el subespacio por un escalar c , el producto cv está en el subespacio.

Nótese que el vector cero debe pertenecer a todo subespacio ya que se puede dar el caso de $c = 0$ y entonces por la regla 2, $0v = 0$ debe estar en el subespacio¹.

Nótese también que no todo subconjunto de un espacio vectorial es un subespacio.

El caso más extremo para un espacio o subespacio es cuando sólo contiene un vector, el vector cero. Es un espacio de “dimensión cero” y geométricamente puede interpretarse como un único punto en el origen.

Ejemplo: matrices de 2x2 en las que las sumas de filas y de columnas son todas iguales. Base:

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\}.$$

¹Recuerde no confundir el escalar 0 con el vector 0

CAPÍTULO 4

TRANSFORMACIONES LINEALES

Sean V y W dos espacios vectoriales (posiblemente iguales). Una transformación lineal T es una función $T : V \longrightarrow W$ que cumple con las siguientes propiedades:

1. $T(\alpha x) = \alpha T(x)$
2. $T(x + y) = T(x) + T(y)$

Siendo x y y vectores del espacio V y α un escalar.

Ejemplos de transformación lineal:

1. Derivación de funciones reales continuas en un intervalo.
2. Integración de funciones reales continuas en un intervalo.
3. Derivación de polinomios de grado n (pasa vectores en el espacio P_n de los polinomios de grado n al espacio P_{n-1}).
4. La multiplicación de un polinomio por otro polinomio fijo. Por ejemplo, multiplicar polinomios de grado n por el polinomio $3 + 4x$ es una transformación lineal que pasa vectores de P_n a P_{n+1} .
5. Rotación de vectores en \mathbb{R}^n .
6. Reflexión de vectores en \mathbb{R}^n .
7. Proyección de vectores en \mathbb{R}^n .

La multiplicación de una matriz por un vector es una transformación lineal, ya que se verifican las propiedades requeridas: $A(\alpha x) = \alpha Ax$ y $A(x + y) = Ax + Ay$.

Si se conoce una base $\{v_1, v_2, \dots, v_n\}$ del espacio V y una base $\{w_1, w_2, \dots, w_m\}$ del espacio W , entonces toda transformación lineal de V en W se puede representar como una multiplicación de una matriz por un vector.

Suponga que la transformación lineal T transforma el vector x del espacio V en el vector $T(x)$ del espacio W . Entonces, cada vector v_j de la base de V tiene un transformado z_j en W y, por definición, cada uno de estos z_j puede ser expresado como combinación lineal de la base de W . Tenemos entonces

$$(4.0.1) \quad T(v_j) = z_j = \sum_{i=1}^m \lambda_{ij} w_i$$

.

Suponga que se tiene un vector x del espacio V cuyas coordenadas con respecto a la base conocida son $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$, entonces tenemos que

$$x = \sum_{j=1}^n \alpha_j v_j$$

y al aplicar la transformación lineal T sobre x nos queda

$$\begin{aligned}
 T(x) &= T\left(\sum_{j=1}^n \alpha_j v_j\right) \\
 &= \sum_{j=1}^n T(\alpha_j v_j) \\
 &= \sum_{j=1}^n \alpha_j T(v_j) \\
 &= \sum_{j=1}^n \alpha_j \sum_{i=1}^m \lambda_{ij} w_i \\
 &= \sum_{i=1}^m \left(\sum_{j=1}^n \lambda_{ij} \alpha_j\right) w_i \\
 &= \sum_{i=1}^m \beta_i w_i
 \end{aligned}$$

Nótese que $(\beta_1, \beta_2, \dots, \beta_m)$ son las coordenadas de $T(x)$ con respecto a la base de W , las cuales resultan de $\beta_i = \sum_{j=1}^n \lambda_{ij} \alpha_j$, el cual se puede representar matricialmente de la siguiente manera:

$$T(x) = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m1} & \lambda_{m2} & \cdots & \lambda_{mn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix}$$

Toda transformación lineal de \mathbb{R}^n a \mathbb{R}^m puede ser representada como una matriz de $m \times n$ y toda matriz de $m \times n$ es una transformación lineal del espacio \mathbb{R}^n a \mathbb{R}^m . Es decir, el concepto de transformación lineal y matriz es uno y el mismo.

CAPÍTULO 5

ESPACIOS FUNDAMENTALES

Sea una Matriz

$$A = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 5 \\ -1 & -3 & 3 & 0 \end{bmatrix}$$

y su matriz escalonada asociada

$$U = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

y la matriz L subdiagonal tal que

$$A = LU$$

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 2 & 1 \end{bmatrix}$$

Los cuatro espacios fundamentales están asociados a las transformaciones Ax y $(A^T)y$, donde A es $m \times n$.

- *Espacio de columnas* de A , $R(A)$
- *Espacio de las filas* de A , $R(A^T)$
- *Espacio Nulo* de A , $N(A)$
- *Espacio Nulo a la izquierda* de A , $N(A^T)$

$N(A)$ y $R(A^T)$ son subespacios de \mathbb{R}^n ; $R(A)$ y $N(A^T)$ son subespacios de \mathbb{R}^m .

En las explicaciones sucesivas se asumirá que A es una matriz $m \times n$ y su matriz escalonada asociada U tiene r pivotes. Se asumirá también que A es una transformación que pasa del espacio \mathbb{R}^n al espacio \mathbb{R}^m .

5.1. Espacio de las filas de A , $R(A^T)$

Este es el espacio de todas las combinaciones lineales de los vectores fila de A , o lo que es lo mismo, las columnas de A^T . Este espacio es el mismo generado por las filas de U y tiene por dimensión el número de pivotes r . Las filas con el pivote no nulo son independientes y forman una base del espacio.

El espacio generado por las filas de U es el mismo que el generado por las filas de A debido a que las filas de U son combinaciones lineales de las filas de A .

En el ejemplo, los vectores $\begin{bmatrix} 1 \\ 3 \\ 3 \\ 2 \end{bmatrix}$ y $\begin{bmatrix} 0 \\ 0 \\ 3 \\ 1 \end{bmatrix}$ constituyen una base para $R(A^T)$. Cualquier vector b

que hace que el sistema $A^T y = b$ sea compatible es combinación lineal de $\begin{bmatrix} 1 \\ 3 \\ 3 \\ 2 \end{bmatrix}$ y $\begin{bmatrix} 0 \\ 0 \\ 3 \\ 1 \end{bmatrix}$.

Este espacio es un subespacio de \mathbb{R}^n ya que está conformado por vectores que están en \mathbb{R}^n .

5.2. Espacio Nulo de A , $N(A)$

$$N(A) = \{x : Ax = 0\}$$

Las soluciones de $Ax = 0$ son las mismas de $Ux = 0$.

Se obtienen los vectores de la base hallando la solución de $Ux = 0$. Esto se hace fácilmente dándole a una variable libre el valor 1 y a las restantes libres el valor 0. Como hay $n - r$ variables libres y una solución para cada una, y como estas soluciones son independientes, la dimensión del espacio es $n - r$.

En el caso de la matriz U del ejemplo, las variables asociadas a las columnas 1 y 3 son básicas y las correspondientes a las columnas 2 y 4 son libres. De la matriz U se tiene:

$$x_3 = -\frac{1}{3}x_4$$

$$x_1 = -2x_4 - 3x_3 - 3x_2 = -x_4 - 3x_2$$

Entonces la solución general de $Ux = 0$ es

$$x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ -\frac{1}{3} \\ 1 \end{bmatrix}$$

Los vectores $\begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ y $\begin{bmatrix} -1 \\ 0 \\ -\frac{1}{3} \\ 1 \end{bmatrix}$ son una base del espacio nulo. Como se ve, cualquier vector x combinación lineal de estos vectores (asignándoles valores cualesquiera a las variables libres) hace que Ax produzca el vector $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

Otra manera rápida de encontrar estos vectores es asignar a cada variable libre, una a una, el valor 1 y a las otras el valor 0.

Este espacio es un subespacio de \mathbb{R}^n .

5.3. Espacio de las columnas de A , $R(A)$

Este es el espacio de las combinaciones lineales de las columnas de A . Representa el conjunto de vectores b para los cuales el sistema $Ax = b$ es compatible.

Los r pivotes de U indican las columnas independientes. El espacio de las columnas de A no es el mismo que el de U , pero las columnas de A que coinciden con las que tienen pivotes en U son independientes y forman una base de $R(A)$.

La razón de esto es que $Ax = 0$ tiene exactamente las mismas soluciones que $Ux = 0$. $Ax = 0$ expresa una dependencia lineal entre las columnas de A dada por los coeficientes de x . Esta dependencia es igual a la que existe entre las columnas de U al resolver $Ux = 0$. Por lo tanto, si un conjunto de columnas de U es independiente, entonces también lo son las columnas correspondientes en A y viceversa.

En el ejemplo, puede verse que la segunda columna tanto de U como de A equivale a tres veces la primera y que la cuarta equivale a la suma de la primera columna más un tercio de la tercera. Los

vectores $\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$ y $\begin{bmatrix} 3 \\ 9 \\ 3 \end{bmatrix}$ son una base del espacio de las columnas de A y cualquier vector b para

los cuales el sistema $Ax = b$ es compatible puede ser representado como una combinación lineal de ambos.

Se obtiene entonces el siguiente resultado fundamental: La dimensión del espacio de filas de A y del espacio de columnas es la misma, r , y los pivotes de U indican una submatriz cuadrada de r filas y columnas que forman una matriz invertible. Esta dimensión r es el Rango de A .

En el caso del ejemplo, los pivotes determinan la submatriz de A $\begin{bmatrix} 1 & 3 \\ 2 & 9 \end{bmatrix}$ que es invertible, pues al escalarizar quedan dos pivotes.

Este espacio es un subespacio de \mathbb{R}^m y también es conocido como recorrido de A ya que incluye a todos los vectores en el recorrido de la función Ax .

5.4. Espacio Nulo a la izquierda de A , $N(A^T)$

Este es el espacio nulo de A^T . Es decir, el espacio de los vectores y tales que $A^T y = 0$, o lo que es lo mismo $y^T A = 0$. Si se tiene $A = LU$ o $PA = LU$ y U tiene sus últimas $m - r$ filas nulas, entonces $U = L^{-1}A$ y las últimas $m - r$ filas de L^{-1} o $L^{-1}P$ constituyen una base del espacio nulo a la izquierda, ya que estas filas multiplicadas por la matriz A producen las filas de ceros en la matriz U .

En el ejemplo,

$$L^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 5 & -2 & 1 \end{bmatrix}, U = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Entonces el vector $\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}$ es una base del espacio nulo de A^T , ya que cualquier vector y múltiplo

de $\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}$ hace que $A^T y = 0 \Leftrightarrow y^T A = 0$.

Este espacio es un subespacio de \mathbb{R}^m .

5.5. Bases complementarias

$R(A)$ y $N(A^T)$ son subespacios de \mathbb{R}^m de dimensión r y $m - r$, ambas bases conforman una base del espacio total. Ídem para el espacio de filas y espacio nulo de A .

Más adelante se probará que son ortogonales.

5.6. Solución de sistemas

Dado el sistema $Ax = b$, que tenga solución es equivalente a que b sea combinación lineal de las columnas de A . Luego, $b \in R(A)$.

Si se define *matriz ampliada del sistema*, $A + b$, a la matriz que se obtiene de agregar la columna b a A , decir que el sistema $Ax = b$ tiene solución es equivalente a decir que: $\text{rango}(A + b) = \text{rango}(A)$.

El proceso de eliminación gaussiana se realiza simultáneamente sobre $A + b$ por lo tanto:

$$Ax = b \Leftrightarrow Ux = L^{-1}b$$

$\text{Rango}(A) = r$. El vector $L^{-1}b$ debe tener sus últimos $m - r$ valores nulos para que el sistema sea compatible.

Suponga que queremos resolver el sistema $Ax = b$, con la matriz A del ejemplo anterior y el vector

$$b = \begin{bmatrix} 2 \\ 1 \\ -8 \end{bmatrix}. \text{ El vector } L^{-1}b \text{ es entonces } \begin{bmatrix} 2 \\ -3 \\ 0 \end{bmatrix}. \text{ La matriz ampliada } A+b \text{ es } \begin{bmatrix} 1 & 3 & 3 & 2 & 2 \\ 2 & 6 & 9 & 5 & 1 \\ -1 & -3 & 3 & 0 & -8 \end{bmatrix},$$

la cual al aplicar el proceso de eliminación queda $\begin{bmatrix} 1 & 3 & 3 & 2 & 2 \\ 0 & 0 & 3 & 1 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$. Puede observarse que el rango de la matriz ampliada es 2, al igual que la matriz A , por lo tanto $Ax = b$ tiene solución para

$$b = \begin{bmatrix} 2 \\ 1 \\ -8 \end{bmatrix}.$$

Si pasamos de la notación matricial a la notación de ecuaciones, nos quedan las ecuaciones

$$\begin{aligned} x_1 + 3x_2 + 3x_3 + 2x_4 &= 2 \\ 3x_3 + x_4 &= -3 \end{aligned}$$

Está claro que las variables x_1 y x_3 , que están en las posiciones de pivote son básicas y x_2 y x_4 son libres. Al resolver hacia atrás nos queda:

$$\begin{aligned} x_3 &= -1 - \frac{1}{3}x_4 \\ x_1 &= 5 - 3x_2 - x_4 \end{aligned}$$

El vector solución x puede ser reescrito como:

$$\begin{bmatrix} 5 - 3x_2 - x_4 \\ x_2 \\ -1 - \frac{1}{3}x_4 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \\ -1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ -\frac{1}{3} \\ 1 \end{bmatrix}$$

Obsérvese que la solución de este sistema se parece bastante a la solución del sistema homogéneo

$Ax = 0$ (espacio nulo de A). La única diferencia es que se añade el vector $\begin{bmatrix} 5 \\ 0 \\ -1 \\ 0 \end{bmatrix}$. Este vector es

lo que se conoce como una solución particular de $Ax = b$. La solución particular proporciona una solución del sistema y los dos términos adicionales agregan aún más soluciones, ya que satisfacen $Ax = 0$.

La solución general x de un sistema $Ax = b$ siempre es la suma de una solución particular x_p más la solución del sistema homogéneo ya que si $Ax = b$ y $Ax_p = b$ entonces $Ax - Ax_p = b - b = 0$, luego $x - x_p \in N(A)$. Entonces $x - x_p = x_h$, siendo x_h la solución general del sistema homogéneo, es decir el conjunto de todas las combinaciones lineales de los vectores de la base de $N(A)$ y por lo tanto $x = x_p + x_h$.

Una manera rápida de encontrar la solución particular es asignar 0 a todas las variables libres y resolver el sistema $U_p x_p = L^{-1}b$, siendo U_p la submatriz de U que incluye únicamente las colum-

nas con las variables básicas. En el ejemplo $U_p = \begin{bmatrix} 1 & 3 \\ 0 & 3 \\ 0 & 0 \end{bmatrix}$, $L^{-1}b = \begin{bmatrix} 2 \\ -3 \\ 0 \end{bmatrix}$ y nos quedarían únicamente las ecuaciones

$$\begin{aligned} x_1 + 3x_3 &= 2 \\ 3x_3 &= -3 \end{aligned}$$

De donde $x_3 = -1$ y $x_1 = 5$. De ahí que la solución particular sea $\begin{bmatrix} 5 \\ 0 \\ -1 \\ 0 \end{bmatrix}$.

5.6.1. Matrices de rango completo. Una matriz A de m filas y n columnas es de rango completo si $Rango(A) = \min(m, n)$. En el caso de matrices de rango completo se presentan dos casos con respecto a las soluciones del sistema $Ax = b$:

EXISTENCIA: $Ax = b$ es soluble y tiene *al menos* una solución x para todo b si y sólo si las columnas generan \mathbb{R}^m , es decir $rango(A) = m$. En este caso existe una matriz inversa a la derecha C tal que $AC = I_m$. Esto es posible sólo si $m \leq n$ (más columnas que filas o igual número).

UNICIDAD: $Ax = b$ tiene *a lo sumo* una solución x para todo b si y solo si las columnas son linealmente independientes, es decir $Rango(A) = n$. En este caso existe una matriz inversa a la izquierda B tal que $BA = I_n$. Esto es posible sólo si $m \geq n$ (más filas que columnas o igual número).

En el primer caso, una posible solución es $x = Cb$, puesto que $Ax = ACb = b$. Pero habrá más soluciones si existen otras inversas a la derecha.

En el segundo caso, si hay una solución a $Ax = b$, entonces ésta tiene que ser $x = BAx = Bb$, pero podría no haber solución.

Si $n > m$ entonces una inversa a la derecha de A es $A^T(AA^T)^{-1}$.

Si $m > n$ entonces una inversa a la izquierda de A es $(A^T A)^{-1} A^T$.

Si $m = n$ la inversa a la izquierda es la misma que la de la derecha y es A^{-1} .

En general, si las columnas de A son linealmente independientes ($\text{Rango} = n$) $A^T A$ es cuadrada, simétrica e invertible.

$A^T A$ obviamente es una matriz cuadrada de $n \times n$. Es simétrica porque $(A^T A)^T = A^T A^{TT} = A^T A$.

Para que $A^T A$ sea invertible es necesario que sus columnas (y por ende sus filas) sean linealmente independientes. Es decir, debe ser de rango n . Esto equivale a decir que su espacio nulo tenga dimensión cero.

El espacio nulo de $A^T A$ siempre es igual al espacio nulo de A . Para demostrar esto hay que demostrar que los vectores del espacio nulo de A están en el espacio nulo de $A^T A$ y viceversa.

Lo primero se verifica al ver que si un vector x está en $N(A)$, entonces $Ax = 0$, y $A^T Ax = 0$; por lo tanto x está en $N(A^T A)$.

Para demostrar lo segundo, supongamos que $x \in N(A^T A)$, entonces $(A^T A)x = 0$, entonces $x^T A^T A x = 0$ y $(Ax)^T A x = 0$. Para demostrar que $(Ax)^T A x$ es cero, multipliquemos a la derecha en ambos lados de la ecuación por x y nos queda $(Ax)^T (Ax) = 0$. Esto es el producto de un vector fila

por el equivalente vector columna. Si $Ax = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix}$, entonces $(Ax)^T (Ax)$ será $\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_m^2$,

lo cual es una suma de números no negativos (ver Sección 6.1). La única manera de que esta suma sea cero es que todos los α_i sean cero y por lo tanto tanto $Ax = 0$; entonces $x \in N(A)$.

Si las columnas de A son linealmente independientes, entonces el espacio nulo de A tiene dimensión cero y por consiguiente el espacio nulo de $A^T A$ tiene también dimensión cero y por lo tanto $A^T A$ es invertible. Por esta razón $(A^T A)^{-1} A^T$ es una inversa a la izquierda de A cuando $\text{Rango}(A) = n$ y la solución de $Ax = b$ es $x = (A^T A)^{-1} A^T b$.

Si las filas de A son linealmente independientes, el mismo análisis aplica para determinar que el espacio nulo de A^T tiene dimensión cero y por lo tanto AA^T es invertible. Entonces $A^T (AA^T)^{-1}$ es una inversa a la derecha cuando $\text{Rango}(A) = m$. En este caso la solución de $Ax = b$ es $x = A^T (A^T A)^{-1} b$.

CAPÍTULO 6

ORTOGONALIDAD

6.1. Producto escalar

DEFINICIÓN. El producto escalar de dos vectores x y y ($x, y \in \mathbb{R}^n$) está definido como

$$(6.1.1) \quad (x, y) = x^T y = \sum_{j=1}^n x_j y_j$$

El producto escalar permite introducir los conceptos de longitud de un vector y distancia entre dos vectores.

DEFINICIÓN. La longitud o norma $\|x\|$ de un vector x en \mathbb{R}^n es la raíz cuadrada positiva de

$$(6.1.2) \quad \|x\|^2 = x_1^2 + x_2^2 + \cdots + x_n^2 = x^T x$$

Esto equivale a aplicar la fórmula del teorema de Pitágoras $n - 1$ veces, añadiendo una dimensión en cada paso.

DEFINICIÓN. La distancia entre dos vectores x y y es $d(x, y) = \|x - y\|$. Esta distancia es conocida como distancia euclídea.

6.2. Algunas Propiedades

$$(x + y)^T (u + v) = x^T u + x^T v + y^T u + y^T v$$

$$x^T x \geq 0, x^T x \text{ sólo es cero cuando } x \text{ es cero.}$$

$$x^T y = y^T x$$

$$z^T (\alpha x + \beta y) = \alpha z^T x + \beta z^T y, \alpha \text{ y } \beta \text{ son escalares.}$$

6.3. Vectores ortogonales

DEFINICIÓN. Dos vectores x y y son ortogonales si $x^T y = 0$.

Si x y y en \mathbb{R}^n son ortogonales, entonces forman un triángulo rectángulo donde las longitudes de los catetos son $\|x\|$, $\|y\|$ y la de la hipotenusa es $\|x - y\|$. Entonces $\|x\|^2 + \|y\|^2 = \|x - y\|^2$.

Aplicando la fórmula 6.1.2, nos queda

$$(x_1^2 + x_2^2 + \cdots + x_n^2) + (y_1^2 + y_2^2 + \cdots + y_n^2) = (x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2$$

El lado derecho queda

$$(x_1^2 + x_2^2 + \cdots + x_n^2) - 2(x_1 y_1 + x_2 y_2 + \cdots + x_n y_n) + (y_1^2 + y_2^2 + \cdots + y_n^2)$$

El lado izquierdo se cancela con parte del derecho y queda

$$0 = -2(x_1y_1 + x_2y_2 + \cdots + x_ny_n)$$

Lo que equivale a

$$x_1y_1 + x_2y_2 + \cdots + x_ny_n = x^T y = 0$$

COROLARIO. El vector cero es ortogonal a todos ya que $x^T 0 = 0^T x = 0$.

COROLARIO. Si $\{x_1, \dots, x_p\}$ son ortogonales y distintos de cero entonces son linealmente independientes.

DEFINICIÓN. U y W subespacios ortogonales si $\forall x \in U, y \in W (x^T y = 0)$.

6.4. Espacios fundamentales

TEOREMA. $R(A)$ y $N(A^T)$ son subespacios ortogonales de \mathbb{R}^m de dimensión r y $m - r$.

DEMOSTRACIÓN. $b \in R(A), c \in N(A^T) \Rightarrow b = Ax, c^T A = 0; \Rightarrow c^T b = c^T Ax = 0x = 0$ □

es decir c y b son ortogonales.

De igual manera se prueba que el espacio nulo de A es ortogonal al espacio de las filas de A .

Otra manera de realizar la demostración es observar que si x pertenece al espacio nulo, entonces $Ax = 0$ y por lo tanto

$$Ax = \begin{bmatrix} \text{fila 1} \\ \text{fila 2} \\ \vdots \\ \text{fila } m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

La fila 1 es ortogonal a x ya que su producto escalar es cero. Lo mismo aplica para todas las filas. Como el vector x es ortogonal a cualquier fila, entonces es ortogonal a cualquier combinación de ellas.

6.5. Ángulos entre vectores

Sean a y b dos vectores en \mathbb{R}^2 tales que sus valores son (a_1, a_2) y (b_1, b_2) y forman ángulos α y β con respecto al eje x . Estos vectores forman un triángulo donde la longitud del tercer lado es $\|b - a\|$. El ángulo θ opuesto al lado $\|b - a\|$ es $\beta - \alpha$.

Recordando que el vector a forma un triángulo rectángulo de hipotenusa $\|a\|$ y catetos a_1 y a_2 con respecto al eje x , se sabe que

$$\sin \alpha = \frac{a_2}{\|a\|} \text{ y } \cos \alpha = \frac{a_1}{\|a\|}$$

Lo mismo aplica para b y su ángulo correspondiente β : el seno es $b_2 / \|b\|$ y el coseno $b_1 / \|b\|$. Puesto que $\theta = \beta - \alpha$, su coseno se puede calcular con la identidad:

$$(6.5.1) \quad \cos \theta = \cos \beta \cos \alpha + \sin \beta \sin \alpha = \frac{a_1 b_1 + a_2 b_2}{\|a\| \|b\|}$$

El numerador en 6.5.1 es $a^T b$ y llegamos a la conclusión de que el coseno del ángulo entre dos vectores a y b es

$$(6.5.2) \quad \cos \theta = \frac{a^T b}{\|a\| \|b\|}$$

6.6. Proyecciones sobre líneas

La proyección del vector b sobre el vector a es el punto p sobre a que está más cercano a b . p está en la misma dirección de a y por lo tanto $p = \lambda a$. La línea de b a p equivale al vector $b - p = b - \lambda a$. Para determinar p basta saber que la línea de b a p debe ser perpendicular a a :

$$a^T(b - \lambda a) = 0, \quad \lambda = \frac{a^T b}{a^T a}$$

Entonces la proyección p del vector b sobre el vector a está dada por

$$p = \lambda a = \frac{a^T b}{a^T a} a = a \frac{a^T b}{a^T a} = \frac{a a^T}{a^T a} b = P b$$

P es una matriz cuadrada que resulta de multiplicar una columna por una fila (aa^T), dividida por el escalar $a^T a$.

La matriz tiene dos propiedades de todas las matrices de proyección:

1. P es simétrica, $P^T = \frac{(aa^T)^T}{a^T a} = \frac{aa^T}{a^T a} = P$
2. $P^2 = P$, $P^2 = \frac{aa^T}{a^T a} \frac{aa^T}{a^T a} = \frac{aa^T aa^T}{a^T aa^T a} = \frac{a(a^T a)a^T}{a^T aa^T a} = \frac{aa^T}{a^T a} = P$

La matriz P tiene rango 1, ya que todas las filas son múltiplos de a , al igual que las columnas.

Ejercicio: ¿Es P invertible?

Ejemplo: La matriz que permite proyectar sobre la línea del vector $a = (1, 1, 1)$ es

$$P = \frac{aa^T}{a^T a} = \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

El rango P es uno. El espacio de las columnas es la línea que pasa por $a = (1, 1, 1)$ y el espacio nulo es el plano perpendicular a a .

En general, como todas las columnas de P son múltiplos de a , Pb está en la línea que pasa sobre a .

Los vectores que satisfacen $a^T b = 0$ son perpendiculares a a y su componente sobre la línea de a es cero. Estos vectores están en el plano perpendicular a a , que es el espacio nulo.

6.7. Proyecciones sobre espacios y aproximaciones con mínimos cuadrados

Cuando se presenta un problema con varias ecuaciones y una sola incógnita el sistema tiene no más de una solución.

$$\begin{aligned} a_1 x &= b_1 \\ a_2 x &= b_2 \\ &\vdots \\ a_m x &= b_m \end{aligned}$$

Si las ecuaciones no son todas equivalentes el sistema no tiene solución. Es decir, sólo hay solución si $b = (b_1, b_2, \dots, b_m)$ es múltiplo de $a = (a_1, a_2, \dots, a_m)$.

Es frecuente que se presenten sistemas incompatibles, y aún así, éstos deben ser resueltos. En estos casos, la opción apropiada es escoger un x tal que minimice el error promedio de las m ecuaciones. La manera más conveniente de definir ese promedio es la suma de cuadrados:

$$E^2 = (a_1x - b_1)^2 + (a_2x - b_2)^2 + \cdots + (a_mx - b_m)^2$$

Si existe una solución exacta a $ax = b$, el error mínimo es $E = 0$. En el caso en que b no es proporcional a a , E^2 será una parábola con el mínimo en el punto donde la derivada del cuadrado del error sea cero:

$$\frac{dE^2}{dx} = 2[(a_1x - b_1)a_1 + (a_2x - b_2)a_2 + \cdots + (a_mx - b_m)a_m] = 0$$

Despejando x nos queda que el punto donde $x = 0$ es

$$\bar{x} = \frac{a_1b_1 + a_2b_2 + \cdots + a_mb_m}{a_1^2 + a_2^2 + \cdots + a_m^2} = \frac{a^T b}{a^T a}$$

6.7.1. Mínimos cuadrados con varias variables. Supóngase que tenemos $Ax = b$, siendo A una matriz $m \times n$. Es decir, que ahora hay n variables en lugar de una. Suponemos que hay más observaciones que variables ($m > n$), así que puede esperarse que $Ax = b$ sea incompatible. Probablemente el vector b no sea una combinación de las columnas de A , es decir, estará fuera del espacio de las columnas.

El problema es encontrar un \bar{x} que minimice el error utilizando el criterio de mínimos cuadrados. $A\bar{x}$ siempre será un vector en el espacio de las columnas, entonces el problema es encontrar el punto $p = A\bar{x}$ que es más cercano a b entre los vectores del espacio de las columnas. Geométricamente se sabe que el vector del error ($b - A\bar{x}$) debe ser perpendicular al espacio de las columnas.

Los vectores perpendiculares al espacio de las columnas están en el espacio nulo a la izquierda. Así, el vector de error $b - A\bar{x}$ debe estar en el espacio nulo de A^T :

$$A^T(b - A\bar{x}) = 0 \Leftrightarrow A^T A\bar{x} = A^T b$$

Si las columnas de A son linealmente independientes, entonces $A^T A$ es invertible y

$$(6.7.1) \quad \bar{x} = (A^T A)^{-1} A^T b$$

La proyección de b en el espacio de las columnas es, por lo tanto,

$$(6.7.2) \quad p = A\bar{x} = A(A^T A)^{-1} A^T b$$

Propiedades:

1. Supóngase que b está en el espacio de las columnas de A , es decir $b = Ax$. Entonces la proyección de b sigue siendo b : $p = A(A^T A)^{-1} A^T Ax = Ax = b$.
2. Supóngase que b es perpendicular al espacio de las columnas de A , es decir $A^T b = 0$. En este caso la proyección es el vector cero: $p = A(A^T A)^{-1} A^T b = A(A^T A)^{-1} 0 = 0$.
3. Cuando A es cuadrada e invertible, el espacio de las columnas es todo el espacio. Todo vector se proyecta en sí mismo: $p = A(A^T A)^{-1} A^T b = AA^{-1}(A^T)^{-1} A^T b = b$.
4. Si A tiene sólo una columna con el vector a la matriz $A^T A$ es el número $a^T a$ y $\bar{x} = a^T b / (a^T a)$.

De la ecuación 6.7.2 observamos que el punto más cercano a b es $p = A(A^T A)^{-1} A^T b$. La matriz

$$P = A(A^T A)^{-1} A^T$$

es una matriz de proyección y tiene dos propiedades básicas:

1. Idempotencia: $P^2 = P$, $P^2 = A(A^T A)^{-1} A^T A(A^T A)^{-1} A^T = A(A^T A)^{-1} A^T = P$

$$2. \text{ Simetría: } P^T = P, P^T = (A(A^T A)^{-1} A^T)^T = A((A^T A)^{-1})^T A^T = A((A^T A)^T)^{-1} A^T = A((A^T A))^{-1} A^T = P$$

DEFINICIÓN. Una matriz de proyección es aquella que es simétrica e idempotente.

Si A es una matriz cuadrada invertible, entonces la matriz de proyección es la identidad:

$$P = A(A^T A)^{-1} A^T = AA^{-1}(A^T)^{-1} A^T = I$$

6.8. Bases ortogonales

DEFINICIÓN. La matriz Q es ortogonal si es cuadrada y sus columnas son ortonormales. Tiene las siguientes propiedades:

- $Q^T = Q^{-1}$
- $\|Qx\| = \|x\|$ preserva longitudes: $(Qx)^T(Qx) = x^T Q^T Qx = x^T x$
- $(Qx)^T(Qy) = x^T y$ preserva ángulos (cosenos de ángulos)
- $b = \sum \beta_j q_j$ entonces $\beta_j = b^T q_j$
- Q^T también es ortogonal

Ejemplos de matrices ortogonales:

$$Q = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}, Q^T = Q^{-1} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$$

Cualquier matriz de permutación es ortogonal.

6.8.1. Matrices rectangulares con columnas ortonormales. Para resolver el problema $Qx = b$ cuando Q es rectangular (más filas que columnas, que es el caso más frecuente). Se tienen n vectores ortonormales q_i que son las columnas de Q . Cada vector tiene m ($m > n$) componentes. $Qx = b$ se resuelve utilizando mínimos cuadrados.

La clave aquí es que $Q^T Q = I_n$. Q^T ya no es la inversa de Q (Q no es cuadrada), pero sigue siendo la inversa a la izquierda. Esto es, $Q^T Q = I_n$, pero $Q Q^T$ es una matriz $m \times m$ que no es I_m .

La matriz de proyección, que normalmente sería $Q(Q^T Q)^{-1} Q^T$, se convierte en $P = Q Q^T$.

6.8.2. Bases ortonormales.

TEOREMA. Supóngase que q_1, q_2, \dots, q_n es una base ortonormal de un espacio. Todo vector b en dicho espacio es la suma de sus proyecciones unidimensionales sobre las líneas de los q 's.

DEMOSTRACIÓN. Todo b es una combinación lineal de los vectores de la base:

$$b = \lambda_1 q_1 + \lambda_2 q_2 + \dots + \lambda_n q_n$$

Si se multiplican ambos lados por q_1^T , entonces queda

$$q_1^T b = \lambda_1 q_1^T q_1$$

A la derecha todos los componentes desaparecen porque $q_1^T q_j = 0$ para $j \neq 1$. Como $q_1^T q_1 = 1$, queda

$$\lambda_1 = q_1^T b$$

El mismo análisis aplica para los otros q_j . Entonces, cualquier vector b se puede expresar como:

$$b = (q_1^T b)q_1 + (q_2^T b)q_2 + \dots + (q_n^T b)q_n$$

Nota: La expresión $\lambda_1 = q_1^T b / q_1^T q_1$ es la proyección sobre una línea. En el caso ortogonal la expresión se simplifica a $\lambda_1 = q_1^T b$. \square

6.8.3. Gram-Schmidt. Siempre resulta más conveniente trabajar con una base ortonormal. Existe un procedimiento denominado “Gram-Schmidt” que permite transformar un conjunto de vectores $\{a_1, a_2, \dots, a_n\}$ LI en un conjunto de vectores ortonormales $\{q_1, q_2, \dots, q_n\}$ que generan el mismo espacio.

Para conseguir q_1 simplemente se toma el vector unitario en la dirección a_1 . Entonces $q_1 = a_1 / \|a_1\|$.

q_2 debe ser un vector ortogonal a q_1 . Si a a_2 se le resta su proyección sobre q_1 el resultado es la proyección de a_2 sobre una dirección ortogonal a q_1 . La proyección de a_2 sobre q_1 es $(q_1^T a_2)q_1$ (ver Sección 6.8.2). Entonces existe un b_2 ortogonal a q_1 que es $b_2 = a_2 - (q_1^T a_2)q_1$. q_2 sería entonces el vector unitario en la dirección b_2 , $q_2 = b_2 / \|b_2\|$.

q_3 debe ser ortogonal a q_1 y q_2 . Si a a_3 se le restan sus proyecciones sobre q_1 y q_2 el resultado es la proyección de a_3 sobre una dirección ortogonal a q_1 y q_2 . Entonces existe un b_3 ortogonal a q_1 y q_2 tal que $b_3 = a_3 - (q_1^T a_3)q_1 - (q_2^T a_3)q_2$. q_3 es el vector unitario en la dirección b_3 , $q_3 = b_3 / \|b_3\|$.

En general, para obtener cada q_k obtenemos un b_k ortogonal a los q_i previamente hallados restando a a_k sus proyecciones sobre los q_i y luego dividiendo b_k por su norma.

$$b_k = a_k - \sum_{i=1}^{k-1} (q_i^T a_k) q_i$$

$$q_k = \frac{b_k}{\|b_k\|}$$

6.8.4. Factorización QR. Supóngase que se tiene una matriz A de $m \times n$ cuyas columnas son los vectores m -dimensionales LI a_1, a_2, \dots, a_n .

Está claro que se puede aplicar el procedimiento de Gram-Schmidt sobre los vectores a_1, a_2, \dots, a_n para obtener un conjunto de vectores ortonormales q_1, q_2, \dots, q_n .

Se sabe que cada vector a_k puede ser expresado como una combinación lineal de los vectores q_i ya que éstos también constituyen una base.

De la Sección 6.8.2 se sabe que si se usa una base ortonormal, cualquier vector a_k es la suma de las proyecciones sobre los vectores de la base. Entonces

$$a_k = (q_1^T a_k)q_1 + (q_2^T a_k)q_2 + \dots + (q_n^T a_k)q_n$$

De la Sección 6.8.3 se sabe que a_1 está en la dirección de q_1 ; que a_2 está en el plano formado por q_1 y q_2 ; y en general a_k está en el espacio generado por los vectores desde q_1 a q_k .

Entonces

$$\begin{aligned} a_1 &= (q_1^T a_1)q_1 \\ a_2 &= (q_1^T a_2)q_1 + (q_2^T a_2)q_2 \\ a_3 &= (q_1^T a_3)q_1 + (q_2^T a_3)q_2 + (q_3^T a_3)q_3 \\ &\vdots \\ a_n &= \sum_{i=1}^n (q_i^T a_n)q_i \end{aligned}$$

Esto puede escribirse matricialmente como:

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} q_1^T a_1 & q_1^T a_2 & \cdots & q_1^T a_n \\ 0 & q_2^T a_2 & \cdots & q_2^T a_n \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & q_n^T a_n \end{bmatrix} = QR$$

donde A y Q son rectangulares y R es cuadrada supra triangular e invertible.

Aplicado al problema de la resolución por mínimos cuadrados (Sección 6.7.1) :

$$(A^T A)\bar{x} = A^T b$$

queda

$$(R^T Q^T Q R)\bar{x} = (R^T R)\bar{x} = R^T Q^T b \Leftrightarrow R\bar{x} = Q^T b$$

que se resuelve muy fácilmente ya que R es triangular.

CAPÍTULO 7

DETERMINANTES

El determinante de una matriz cuadrada es una medida del volumen del paralelepípedo formado por sus vectores filas o columnas.

DEFINICIÓN. El determinante de A es la suma de los $n!$ productos que se pueden formar eligiendo un elemento por fila y uno por columna de A (dicha elección coincide con una matriz de permutación), cada producto está además multiplicado por el signo de la paridad del número de inversiones o intercambios necesarios para convertir la permutación indicada en la identidad.

$$\det A = \sum_{\sigma} a_{1\sigma(1)} \dots a_{n\sigma(n)} (-1)^{\text{signo}(\sigma)}$$

En el caso particular de matrices de 2×2 , $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$.

7.1. Propiedades

1. Si se multiplica una fila por una constante, $\det A$ queda multiplicado por esa constante. Cada término de la sumatoria que se utiliza para calcular $\det A$ incluye un único elemento de cada fila, entonces todos los elementos de la sumatoria quedan multiplicados por la constante.
2. Si el vector que compone una fila cualquiera es la suma de dos vectores (fila $i = v_1^T + v_2^T$) entonces $\det A = \det A_1 + \det A_2$ donde A_1 tiene todas sus filas iguales a las de A excepto que en la fila i tiene el vector v_1^T . A_2 tiene todas sus filas iguales a las de A excepto que en la fila i tiene el vector v_2^T . Supóngase que cada término de la fila i de A es $a_{ij} = v_{1j} + v_{2j}$, entonces

$$\begin{aligned} \det A &= \sum_{\sigma} a_{1\sigma(1)} \dots (v_{1j} + v_{2j}) \dots a_{n\sigma(n)} (-1)^{\text{signo}(\sigma)} \\ \det A &= \sum_{\sigma} a_{1\sigma(1)} \dots (v_{1j}) \dots a_{n\sigma(n)} (-1)^{\text{signo}(\sigma)} + \sum_{\sigma} a_{1\sigma(1)} \dots (v_{2j}) \dots a_{n\sigma(n)} (-1)^{\text{signo}(\sigma)} \\ \det A &= \det A_1 + \det A_2 \end{aligned}$$

3. Si se intercambian dos filas $\det A$ cambia de signo. Al intercambiar dos filas es evidente que la magnitud no cambia de signo, pero el signo cambia ya que hay un intercambio adicional para convertir la permutación en la identidad.
4. $\det I = 1$. En este caso el único producto en la sumatoria que no produce ceros es el de la diagonal y es 1.
5. Si dos filas son iguales, $\det A = 0$. Si se intercambian las filas iguales está claro que $\det A$ no varía. Por la propiedad 3, el signo cambia, entonces $\det A = -\det A$ y por lo tanto $\det A = 0$.
6. La operación de sumar a una fila un múltiplo de otra deja el determinante sin modificaciones. Supóngase que la fila i está formada por el vector fila f_i y la fila k está formada por el vector fila f_k ($i \neq k$). Sea B una matriz igual a A excepto que la fila i es sustituida por $f_i + cf_k$; por las propiedades 1 y 2, $\det B = \det A + c \det C$, donde C tiene las filas iguales a A excepto que la fila i está compuesta por el vector fila f_k . Por la propiedad 5, $\det C = 0$, ya que las filas i y k de C son iguales.

Esta propiedad se puede extender sin dificultad a la operación de sumar a una fila una combinación lineal de las otras filas.

7. Si A tiene una fila de ceros entonces $\det A = 0$. Todos los elementos de la sumatoria incluyen un elemento de la fila de ceros y por lo tanto todos los términos de dicha sumatoria valen cero.
8. Si A es triangular, $\det A$ es el producto de los elementos de la diagonal. Los elementos de la sumatoria siempre incluyen un elemento de cada fila y cada columna. Salvo por la permutación correspondiente a la diagonal, todas las permutaciones incluyen algún elemento debajo de la diagonal y por lo tanto los términos de la sumatoria asociados a dichas permutaciones valen cero.
9. Si A es singular, $\det A = 0$. En este caso se pueden aplicar pasos de eliminación hasta que quede una fila de ceros sin modificar el determinante (propiedad 6). Por la propiedad 7 el determinante es cero.
10. $\det A = \det A^T$. Los términos que resultan al escoger un elemento de cada fila y cada columna son los mismos que los que resultan al escoger un elemento de cada columna y cada fila.
11. $\det AB = (\det A)(\det B)$. Si D es una matriz diagonal, $\det DB = (\det D)(\det B)$ ya que DB resulta de multiplicar cada fila de B por el elemento de la fila correspondiente en D . Por la propiedad 1 $\det DB$ es igual a $\det B$ multiplicado por todos los elementos de la diagonal de D , pero $\det D$ es precisamente el producto de los elementos de la diagonal (propiedad 8).

Utilizando eliminación Gauss-Jordan se puede reducir la matriz A a una diagonal D sin modificar el determinante (excepto por los cambios de signo al intercambiar filas). Los mismos pasos transforman AB a DB sin modificar el determinante. Entonces

$$\det AB = \det DB = (\det D)(\det B) = (\det A)(\det B)$$

si el número de intercambios es impar, $\det AB = -\det DB = -(\det D)(\det B) = (\det A)(\det B)$

12. $\det A^{-1} = \frac{1}{\det A}$. Según las propiedades 4 y 11, $\det AA^{-1} = (\det A)(\det A^{-1}) = \det I = 1$, entonces $\det A^{-1} = \frac{1}{\det A}$.
13. Si Q es ortogonal, $|\det Q| = 1$. Como Q es ortogonal, $QQ^T = I$; entonces $\det(QQ^T) = \det Q \det Q^T = (\det Q)^2 = 1$.
14. Si A es no singular y admite una descomposición $PA = LDU$ entonces $\det P \det A = \prod \text{pivotes}$. Está claro que $\det PA = \det P \det A$ y $\det(LDU) = \det L \det D \det U$. Como L y U son triangulares con unos en sus diagonales, sus determinantes son 1 y como D es diagonal, su determinante es el producto de los pivotes.

7.2. El determinante como medida de volumen

El determinante de una matriz es una medida del volumen del paralelepípedo formado por sus vectores filas o columnas.

$$|\det A| = |\det(QR)| = |\det Q| |\det R| = \det R = \prod_{i=1}^n a_i^T q_i$$

En el caso de dos dimensiones el valor absoluto del determinante es el área del paralelogramo formado por los vectores a_1 y a_2 . Esta área es el producto de la base del paralelogramo por su altura. Si tomamos como base el vector a_1 , cuya longitud es $\|a_1\| = a_1^T q_1$, entonces la altura viene dada por la proyección de a_2 sobre la perpendicular a a_1 (y q_1 que es colineal), es decir, $(a_2^T q_2)q_2$. Esta proyección tiene longitud $a_2^T q_2$ y por lo tanto el área es $(a_1^T q_1)(a_2^T q_2)$.

El mismo razonamiento aplica para los casos de mayor dimensión.

Nótese que si los vectores son linealmente dependientes el volumen es 0, es decir, si la matriz es singular entonces su determinante es nulo.

7.3. Regla de Cramer

7.3.1. Cofactores.

DEFINICIÓN.

$$\text{cofactor } A_{ij} = (-1)^{i+j} \det M_{i,j}$$

donde $M_{i,j}$ es la matriz deducida de A al suprimirle la fila i y la columna j .

TEOREMA. El determinante de A se expresa en función de los cofactores de una fila:

$$\det A = \sum_i a_{ki} A_{ki}$$

También es válido el desarrollo por columnas.

Si se cambia la fila k por la i entonces:

$$0 = \sum_i a_{hi} A_{ki} \quad h \neq i$$

7.3.2. Aplicaciones. La inversa se deduce del desarrollo $A * A^T \text{cofactores} = (\det A)I$. Entonces $\text{inv}(A) = \frac{A^T \text{cofactores}}{\det A}$

$$A = \begin{bmatrix} 1 & 3 & 4 \\ 3 & 1 & 7 \\ 2 & 3 & 3 \end{bmatrix}$$

$$\det(A) = 25$$

$$\text{Cofactor}(A^T) = \begin{bmatrix} -18 & 3 & 17 \\ 5 & -5 & 5 \\ 7 & 3 & -8 \end{bmatrix}$$

La matriz del numerador, $A^T \text{cofactores}$, se forma con la traspuesta de los cofactores de A.

Regla de Cramer. Dado el sistema $Ax = b$ con A invertible, cada componente de la solución x_j se determina por el cociente de determinantes:

$$x_j = \frac{\det B_j}{\det A}$$

donde B_j se obtiene reemplazando la columna j de A por el vector b .

CAPÍTULO 8

VALORES Y VECTORES PROPIOS

DEFINICIÓN. λ y v son un *valor y vector propio* de la matriz cuadrada A si se satisface: $Av = \lambda v$.

Si se ve la matriz A como una transformación del espacio \mathbb{R}^n en sí mismo, pueden verse los vectores propios como aquellas direcciones en las que el vector transformado es colineal con el original. El valor propio es la relación entre la longitud del vector transformado y el original.

Nótese que cuando se habla de un vector propio, lo que realmente es propio es la dirección del vector. Cualquier vector y colineal con un vector propio v es propio también.

$$Ay = A(cv) = cAv = c\lambda v = \lambda y$$

Si $A = \begin{bmatrix} 4 & -5 \\ 2 & -3 \end{bmatrix} \Rightarrow A - \lambda I = \begin{bmatrix} 4 - \lambda & -5 \\ 2 & -3 - \lambda \end{bmatrix}$ y su determinante se anula para $\lambda = -1, 2$.

La ecuación $Av = \lambda v$ es equivalente a $(A - \lambda I)v = 0$, es decir, v pertenece al espacio nulo de $A - \lambda I$, y λ debe ser elegido para que la matriz resultante sea singular, es decir:

$$\det(A - \lambda I) = 0$$

Al determinante $\det(A - \lambda I) = 0$ se le llama *Polinomio característico* y su grado es n , luego tiene n raíces λ_j en el campo complejo.

Para cada λ_j se determina una base del espacio nulo de $A - \lambda_j I$. Esta base genera el espacio de vectores propios asociados al valor propio λ_j . En general suele haber un solo vector en esa base y se dice que forman un par: el valor propio y su vector propio asociado.

En el ejemplo anterior, para $\lambda = -1$ la matriz $A - (-1)I = \begin{bmatrix} 5 & -5 \\ 2 & -2 \end{bmatrix}$ y $v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Nótese que los valores propios pueden ser complejos. Por ejemplo considérese la matriz $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. En este caso $\det(A - \lambda I) = \lambda^2 + 1$. Este polinomio tiene como raíces $\lambda_1 = i$ y $\lambda_2 = -i$. Esta matriz representa una rotación de 90 grados y por lo tanto no existen valores propios reales.

8.1. Propiedades

8.1.1. Traza e Identidad.

TEOREMA. Los valores propios satisfacen las siguientes identidades:

$$\text{Traza}(A) = \sum_{i=1}^n \lambda_j = \sum_{i=1}^n a_{ii} \quad \text{Det}(A) = \prod_{i=1}^n \lambda_j$$

DEMOSTRACIÓN. En el caso de $\text{Det}(A - I\lambda)$ cualquier término que incluya un a_{ij} fuera de la diagonal excluye los términos $a_{ii} - \lambda$ y $a_{jj} - \lambda$. Por lo tanto dichos términos no incluyen $(-\lambda)^{n-1}$. El coeficiente de $(-\lambda)^{n-1}$ debe provenir de la diagonal principal y es $a_{11} + a_{22} + \dots + a_{nn} = \lambda_1 + \dots + \lambda_n$. Recuerdese que en un polinomio de grado n el coeficiente del factor $n - 1$ es opuesto de la suma de las raíces.

Cuando $\lambda = 0$, $\text{Det}(A - I\lambda) = \text{Det}(A)$. Si $\lambda = 0$ lo que queda del polinomio es el término independiente y éste siempre es igual al producto de las raíces, que en este caso son los valores propios. \square

Ejercicio: Si $\lambda = 0$ es valor propio de A ¿cuál es el subespacio de vectores propios asociado?

8.1.2. Matriz Identidad.

TEOREMA. Si $A = I$ todo vector es propio para el valor propio 1.

DEMOSTRACIÓN. Si se sigue el método anterior se tiene: $\text{Det}(I - \lambda I) = (1 - \lambda)^n$ luego 1 es el único valor propio y el espacio nulo es R^n , es decir una base del espacio nulo de $A - \lambda I$ está formada por cualquier conjunto de n vectores LI de R^n . \square

8.1.3. Matrices de Proyección.

TEOREMA. Si P es una matriz de proyección entonces cualquier vector del espacio de las columnas de P es un vector propio asociado al valor propio 1 y cualquier vector del espacio nulo de P es un vector propio asociado al valor propio 0.

DEMOSTRACIÓN. Una matriz de proyección P produce la proyección de un vector c sobre su espacio de las columnas. Si c está en el espacio de las columnas $Pc = c$. Para cualquier vector d que sea ortogonal al espacio de las columnas (es decir, que esté en el espacio nulo) su proyección es el vector cero y d es un vector propio asociado al valor propio cero. $Pd = 0 = 0d$.

Ningún otro vector es propio ya que P siempre los va a proyectar sobre el espacio de las columnas. \square

8.1.4. Inversas.

TEOREMA. Si A es invertible, A^{-1} tiene los mismos vectores propios y sus valores propios son los inversos de los valores propios de A .

DEMOSTRACIÓN. Sea v un vector propio de la matriz invertible A . Entonces $Av = \lambda v$. Entonces $v = A^{-1}\lambda v$ y $A^{-1}v = \lambda^{-1}v$. \square

8.2. Forma Diagonal

TEOREMA. Si A tiene n vectores propios linealmente independientes entonces es diagonalizable.

DEMOSTRACIÓN. Sea S la matriz cuyas columnas son los vectores propios de A .

$$S = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$$

S es no singular por ser vectores LI, luego existe su inversa S^{-1} . Entonces si Λ es la matriz diagonal formada por los valores propios se cumple:

$$AS = A \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} = \begin{bmatrix} \lambda_1 v_1 & \lambda_2 v_2 & \dots & \lambda_n v_n \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & \dots & 0 \\ 0 & \dots & \dots & 0 \\ 0 & \dots & 0 & \lambda_n \end{bmatrix}$$

$$AS = S\Lambda \quad \text{donde } \Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & \dots & 0 \\ 0 & \dots & \dots & 0 \\ 0 & \dots & 0 & \lambda_n \end{bmatrix}$$

Es decir: $S^{-1}AS = \Lambda$ o $A = S\Lambda S^{-1}$. □

TEOREMA. Si los valores propios $\lambda_1, \lambda_2, \dots, \lambda_n$ son todos distintos entonces $\{v_1, v_2, \dots, v_n\}$ es linealmente independiente.

A continuación se dará una demostración para el caso de dimensión 2.

DEMOSTRACIÓN. Supóngase que

$$(8.2.1) \quad c_1 v_1 + c_2 v_2 = 0$$

con c_1 y c_2 distintos de cero.

Multiplicando la expresión 8.2.1 por A queda

$$(8.2.2) \quad A(c_1 v_1 + c_2 v_2) = c_1 A v_1 + c_2 A v_2 = c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 = 0$$

Restando 8.2.2 menos λ_2 veces 8.2.1 queda

$$(8.2.3) \quad c_1 v_1 (\lambda_1 - \lambda_2) = 0$$

Si c_1 es distinto de cero, entonces $\lambda_1 - \lambda_2$ debe ser cero, lo cual es una contradicción porque por hipótesis λ_1 y λ_2 son diferentes.

El mismo análisis se extiende a dimensiones mayores: se asume que existe una combinación lineal que produce ceros, se multiplica por A , se le resta menos λ_n veces la combinación original y queda una combinación de los vectores v_1, v_2, \dots, v_{n-1} que produce cero. Repitiendo los mismos pasos progresivamente se llega a un múltiplo de v_1 que produce cero (como en 8.2.3). Esto fuerza que c_1 y todos los c_i deben ser cero. □

8.3. Aplicaciones

8.3.1. Potencias de una matriz. Si $A = S\Lambda S^{-1}$ entonces $A^n = S\Lambda^n S^{-1}$.

Por ejemplo, para $n = 2$, $A^2 = S\Lambda S^{-1}S\Lambda S^{-1} = S\Lambda^2 S^{-1}$

8.3.1.1. *Números de Fibonacci.* $F_{n+1} = F_n + F_{n-1}$, $F_0 = 0, F_1 = 1$

Podemos plantear el sistema de ecuaciones:

$$\begin{aligned} F_{n+1} &= F_n + F_{n-1} \\ F_n &= F_{n-1} \end{aligned}$$

Sea $U_n = \begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix}$, entonces $U_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ y $U_{n+1} = \begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix}$. Nos queda

$$U_{n+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} U_n = AU_n$$

$$U_n = AU_{n-1}$$

\vdots

$$U_2 = AU_1$$

$$\text{Entonces } U_{n+1} = A^n U_1.$$

8.3.1.2. *Problema de Población.* Supóngase que de EEUU sale cada año el 20 % de su población e ingresa el 10 % del resto del mundo. Sea y población fuera, y x población dentro.

$$x_{n+1} = 0,8x_n + 0,1y_n$$

$$y_{n+1} = 0,2x_n + 0,9y_n$$

Si $U_n = \begin{bmatrix} x_n \\ y_n \end{bmatrix}$, entonces

$$U_{n+1} = \begin{bmatrix} 0,8 & 0,1 \\ 0,2 & 0,9 \end{bmatrix} U_n$$

$$U_{n+1} = A^n U_1$$

Los valores propios de la matriz A son las raíces de

$$\begin{vmatrix} 0,8 - \lambda & 0,1 \\ 0,2 & 0,9 - \lambda \end{vmatrix} = (0,8 - \lambda)(0,9 - \lambda) - 0,1 \times 0,2 = \lambda^2 - 1,7\lambda + 0,7$$

Las raíces son $\lambda_1 = 1$ y $\lambda_2 = 0,7$. Los vectores propios asociados son $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ y $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$, respectivamente.

La matriz de vectores propios es $S = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}$ y su inversa es $S^{-1} = \frac{1}{3} \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}$. La matriz diagonal de valores propios es $\Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 0,7 \end{bmatrix}$.

$$\text{Como } A^n = S\Lambda^n S^{-1}, U_{n+1} = \frac{1}{3} \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1^n & 0 \\ 0 & 0,7^n \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} U_1.$$

Para valores grandes de n ,

$$U_{n+1} = \frac{1}{3} \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3}(x_1 + y_1) \\ \frac{2}{3}(x_1 + y_1) \end{bmatrix}$$

Nótese que a largo plazo, si se mantiene la tendencia, un tercio de toda la población estará en EEUU y dos tercios en el resto del mundo, sin importar la proporción inicial.

8.4. Matrices simétricas

TEOREMA. Si A es simétrica entonces sus valores propios son reales.

DEMOSTRACIÓN. Sea v un valor propio de A . Entonces

$$(8.4.1) \quad \|Av\|^2 = (Av)^T Av = v^T A^T Av = v^T A^2 v$$

Como v es un valor propio $Av = \lambda v$ y

$$(8.4.2) \quad A^2 v = AAv = A\lambda v = \lambda Av = \lambda^2 v$$

Sustituyendo 8.4.1 en 8.4.2 queda $\|Av\|^2 = v^T \lambda^2 v = \lambda^2 v^T v = \lambda^2 \|v\|^2$, de donde

$$\lambda^2 = \frac{\|Av\|^2}{\|v\|^2}$$

Claramente, λ^2 es un número real positivo. Si λ es un complejo de la forma $a + bi$, entonces $\lambda^2 = a^2 - b^2 - 2abi$. Esta expresión sólo puede ser real positiva cuando $b = 0$ y por lo tanto λ es un número real. \square

TEOREMA. Si los valores propios de una matriz simétrica A son distintos, entonces los vectores propios correspondientes son ortogonales.

DEMOSTRACIÓN. Sean λ_1 y λ_2 dos valores propios cualesquiera de A y sean x y y sus correspondientes vectores propios tales que $Ax = \lambda_1 x$ y $Ay = \lambda_2 y$.

Multiplicando por y^T y x^T queda

$$(8.4.3) \quad y^T Ax = y^T \lambda_1 x = \lambda_1 y^T x$$

y

$$(8.4.4) \quad x^T Ay = x^T \lambda_2 y = \lambda_2 x^T y$$

.

Trasponiendo la última expresión queda

$$(8.4.5) \quad y^T Ax = \lambda_2 y^T x$$

.

Igualando 8.4.3 y 8.4.5 queda $\lambda_1 y^T x = \lambda_2 y^T x$, lo cual equivale a

$$(8.4.6) \quad y^T x (\lambda_1 - \lambda_2) = 0$$

Como λ_1 y λ_2 son diferentes, entonces $y^T x$ debe ser cero y por lo tanto x y y son ortogonales. \square

COROLARIO. Una transformación sobre una base ortogonal de vectores propios se expresa como una dilatación en cada dirección ortogonal

$$A \left(\sum_{k=i}^n \beta_j v_j \right) = \sum_{k=i}^n \beta_j Av_j = \sum_{k=i}^n \beta_j \lambda_j v_j$$

COROLARIO. Si los v_j se toman ortonormales entonces $S^T = S^{-1}$ luego, $A = SAS^T$.

8.5. Formas cuadráticas

DEFINICIÓN. Una forma cuadrática $F(x)$ es una función de varias variables, -de las variables que forman el vector x - que es suma de monomios de segundo grado:

$$F(x) = \sum_{i=1}^n \sum_{j=1}^n a_{i,j} x_i x_j$$

esta función se expresa sintéticamente así:

$$F(x) = x^T A x$$

donde los elementos fuera de la diagonal son la mitad de los coeficientes de $x_i x_j$

EJEMPLO. $F(x, y, z) = 2x^2 + 4xy + y^2 + 3z^2 - 6yz$ entonces:

$$A = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 1 & -3 \\ 0 & -3 & 3 \end{bmatrix}$$

Nótese que A es simétrica.

8.5.1. Gradiente de una forma cuadrática. El gradiente de una función F es el vector de las derivadas parciales de F respecto de las x_j :

$$\nabla F(x) = \begin{bmatrix} \frac{\partial F}{\partial x_1} \\ \frac{\partial F}{\partial x_2} \\ \vdots \\ \frac{\partial F}{\partial x_n} \end{bmatrix}$$

$$F(x) = x^T A x \rightarrow \nabla F(x) = 2Ax$$

Notas:

- para las formas lineales $F(x) = b^T x$ su gradiente es $\nabla F(x) = b$
- para las funciones de segundo grado completas:

$$F(x) = x^T A x + b^T x + c \rightarrow \nabla F(x) = 2Ax + b$$

8.5.2. Positividad de una matriz. La matriz simétrica A es:

- *definida positiva* si $F(x) = x^T A x > 0 \quad \forall x \neq 0$
- *semidefinida positiva* si $F(x) = x^T A x \geq 0 \quad \forall x \neq 0$
- *definida negativa* si $F(x) = x^T A x < 0 \quad \forall x \neq 0$
- *semidefinida negativa* si $F(x) = x^T A x \leq 0 \quad \forall x \neq 0$

8.5.3. Análisis de las formas cuadráticas. Se probará a continuación que el conocimiento de los vectores y valores propios de una matriz simétrica A permite determinar si la forma cuadrática $F(x) = x^T A x$ asociada es de alguno de los tipos anteriormente enunciados.

Previamente, se caracteriza la transformación A y su forma cuadrática asociada usando como base del espacio el conjunto de vectores propios.

Se formará una base ortonormal de vectores propios v_1, \dots, v_n . Cada vector del espacio es una CL de esta base, luego:

Si $x = \sum c_j v_j$ entonces $Ax = \sum c_j A v_j = \sum c_j \lambda_j v_j$, es decir, Ax se obtiene simplemente de la CL de los vectores dilatados por su valor propio asociado. En cuanto a la forma cuadrática:

$$F(x) = x^T A x = \left(\sum c_k v_k^T \right) \left(\sum c_j \lambda_j v_j \right) = \sum c_j^2 \lambda_j$$

Nótese que:

- Si todos los valores propios son positivos $F(x)$ es definida positiva
- Si todos los valores propios son positivos o nulos $F(x)$ es semidefinida positiva
- Si todos los valores propios son negativos $F(x)$ es definida negativa
- Si todos los valores propios son negativos o nulos $F(x)$ es semidefinida negativa
- Si existe al menos un positivo y un negativo $F(x)$ toma valores de distinto signo

Si los vectores propios, v_j , no son de norma 1 el resultado es:

$$F(x) = x^T A x = \left(\sum c_k v_k^T \right) \left(\sum c_j \lambda_j v_j \right) = \sum c_j^2 \lambda_j \|v_j\|^2$$

y no cambia la interpretación de la forma cuadrática ya que su signo sigue dependiendo de los valores propios asociados a la matriz.

8.5.3.1. Curvas de nivel. ¿Cuándo es $F(x) = x^T A x = \text{constante}$? Si se expresa $x^T A x = \sum c_j^2 \lambda_j = \text{constante}$, se determina entonces el punto de corte de la curva de nivel con los ejes definidos por los vectores propios. Si todos los λ_j son positivos la curva de nivel es un elipsoide, es decir una elipse generalizada. Para los vectores sobre el eje v_j , $x = c_j v_j$, $F(x) = c_j^2 \lambda_j = \text{constante}$, entonces: $c_j = \pm \sqrt{\frac{\text{constante}}{\lambda_j}}$. El eje mayor de la elipse corresponde al menor valor propio, ya que al menor λ_j le corresponde el mayor c_j .

EJEMPLO. Sea $F(x) = 2x_1^2 + 4x_1x_2 + 3x_2^2 + 4x_3^2 - 4x_2x_3$ la forma cuadrática que en notación matricial es $F(x) = x^T A x$ con

$$A = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 3 & -2 \\ 0 & -2 & 4 \end{bmatrix}$$

A tiene por matriz de vectores propios

$$S = \begin{bmatrix} -\frac{2}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{3}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix}$$

y valores propios

$$\Lambda = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

luego, la forma cuadrática es semidefinida positiva y vale estrictamente 0 para cualquier vector colineal con

$$v_1 = \begin{bmatrix} -\frac{2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$$

por ejemplo:

$$x = \begin{bmatrix} -4 \\ 4 \\ 2 \end{bmatrix}$$

CAPÍTULO 9

EXTREMOS DE FUNCIONES DE VARIAS VARIABLES

Los candidatos a extremos de una función diferenciable de varias variables deben buscarse entre los puntos estacionarios de la función: es decir, donde su gradiente $\nabla F(x)$ es nulo. Entonces un estudio de extremos se inicia resolviendo el sistema:

$$\nabla F(x) = 0$$

Supóngase que una de las soluciones de la ecuación anterior es x_0 . La nulidad del gradiente no garantiza que F tenga un máximo o un mínimo en el punto x_0 ; luego, es necesario una aproximación de segundo orden en el punto para determinar el comportamiento de la función. Esta aproximación la da el siguiente teorema.

9.1. Teorema de Taylor

$$(9.1.1) \quad F(x) = F(x_0) + (x - x_0)^T \nabla F(x_0) + \frac{1}{2}(x - x_0)^T H(x_0)(x - x_0) + \epsilon \|(x - x_0)\|^2$$

con ϵ un infinitésimo en x_0 .

La condición de estacionaridad, $\nabla F(x_0) = 0$, implica que el comportamiento en el punto depende de la forma cuadrática $(x - x_0)^T H(x_0)(x - x_0)$ que se analiza a través de su positividad.

$H(F(x))$ es el hessiano de $F(x)$ y es una matriz simétrica constituida por todas las segundas derivadas parciales.

$$H(F(x)) = \begin{bmatrix} \frac{\partial^2 F}{\partial x_1^2} & \frac{\partial^2 F}{\partial x_1 \partial x_2} & \frac{\partial^2 F}{\partial x_1 \partial x_3} & \cdots & \frac{\partial^2 F}{\partial x_1 \partial x_n} \\ \frac{\partial^2 F}{\partial x_2 \partial x_1} & \frac{\partial^2 F}{\partial x_2^2} & \frac{\partial^2 F}{\partial x_2 \partial x_3} & \cdots & \frac{\partial^2 F}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial x_n \partial x_1} & \frac{\partial^2 F}{\partial x_n \partial x_2} & \frac{\partial^2 F}{\partial x_n \partial x_3} & \cdots & \frac{\partial^2 F}{\partial x_n \partial x_n} \end{bmatrix}$$

EJEMPLO. $F(x, y) = 7 + 2(x + y)^2 - y \sin y - x^3$ en $(0, 0)$

El gradiente de $F(x, y)$ es $\begin{bmatrix} 4(x + y) - 3x^2 \\ 4(x + y) - \sin y - y \cos y \end{bmatrix}$. En el punto $(0, 0)$ el gradiente es nulo y por lo tanto es un punto estacionario.

El hessiano H es $\begin{bmatrix} 4 - 6x & 4 \\ 4 & 4 - 2 \cos y + y \sin y \end{bmatrix}$. En el punto $(0, 0)$ el hessiano vale $\begin{bmatrix} 4 & 4 \\ 4 & 2 \end{bmatrix}$.

Para calcular los valores propios del hessiano basta resolver $\det(H - I\lambda)$ que es $(4 - \lambda)(2 - \lambda) - 16 = \lambda^2 - 6\lambda - 8$. Este polinomio tiene una raíz positiva y una negativa (recordar que el término independiente es igual al producto de las raíces y al ser negativo es producto de un valor positivo y uno negativo) y por lo tanto la matriz es indefinida y el punto $(0, 0)$ es un punto de silla, como se observa en la Figura 9.1.1.

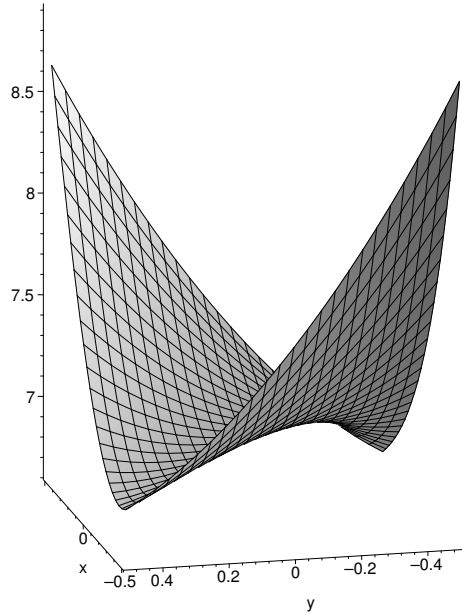


FIGURA 9.1.1. Gráfica de la función $F(x, y) = 7 + 2(x + y)^2 - y \sin y - x^3$

EJEMPLO. $F(x, z) = 2x^2z^2 - 4z - 4x$ en $(1, 1)$

El gradiente es $\begin{bmatrix} 4xz^2 - 4 \\ 4x^2z - 4 \end{bmatrix}$. En $(1, 1)$ el gradiente es nulo y por lo tanto es un punto estacionario.

El hessiano H es $\begin{bmatrix} 4z^2 & 8xz \\ 8xz & 4x^2 \end{bmatrix}$. En $(1, 1)$ el hessiano es $\begin{bmatrix} 4 & 8 \\ 8 & 4 \end{bmatrix}$.

Para calcular los valores propios se resuelve el polinomio característico $(4 - \lambda)^2 - 64 = \lambda^2 - 8\lambda - 48$. En este caso también hay una raíz positiva y una negativa (las raíces son 12 y -4) y por lo tanto $(1, 1)$ es un punto de silla, como se observa en la Figura 9.1.2.

EJEMPLO. $F(x, y) = xe^x \cos y$ en $(-1, 0)$

El gradiente es $\begin{bmatrix} e^x \cos y + xe^x \cos y \\ -xe^x \sin y \end{bmatrix}$. En $(-1, 0)$ el gradiente es nulo y por lo tanto es un punto estacionario.

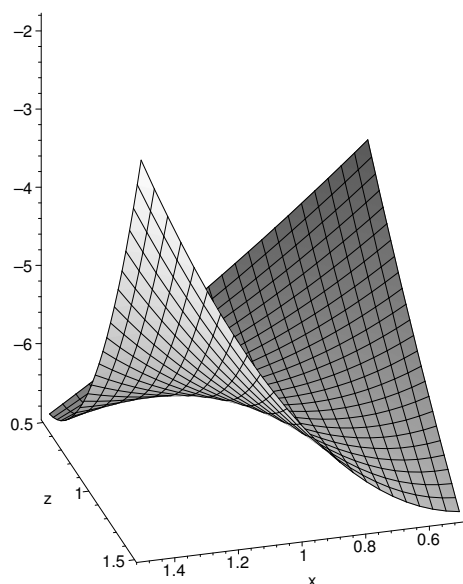
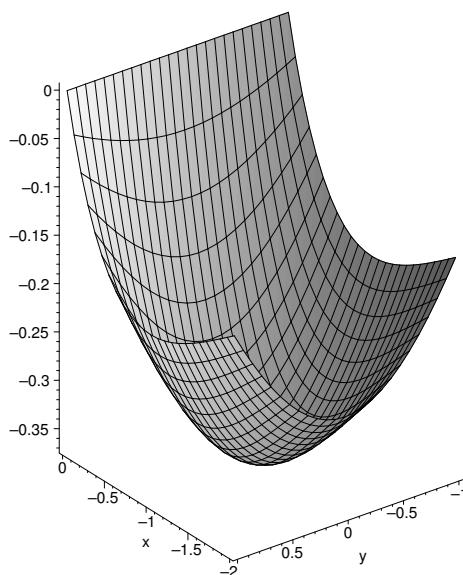
El hessiano es $\begin{bmatrix} 2e^x \cos y + xe^x \cos y & -e^x \sin y - xe^x \sin y \\ -e^x \sin y - xe^x \sin y & -xe^x \cos y \end{bmatrix}$. En $(-1, 0)$ el hessiano es $\begin{bmatrix} e^{-1} & 0 \\ 0 & e \end{bmatrix}$.

Para calcular los valores propios se resuelve el polinomio característico $(e^{-1} - \lambda)(e - \lambda)$. En este caso ambas raíces son positivas y por lo tanto $(-1, 0)$ es un mínimo, como se observa en la Figura 9.1.3.

EJEMPLO. $F(x, y) = (x - y)^2 + x^3$ en $(0, 0)$

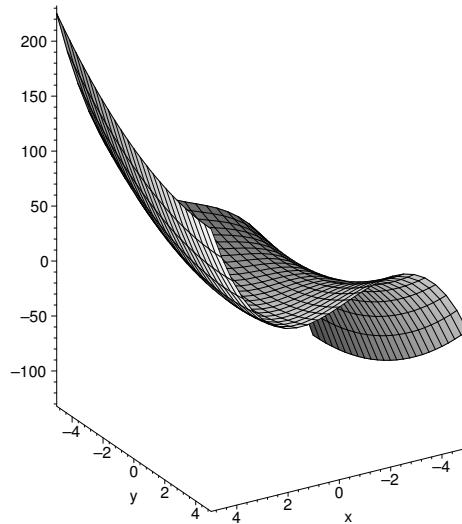
El gradiente es $\begin{bmatrix} 2(x - y) + 3x^2 \\ 2(y - x) \end{bmatrix}$. En $(0, 0)$ el gradiente es nulo y por lo tanto es un punto estacionario.

El hessiano es $\begin{bmatrix} 2 + 6x & -2 \\ -2 & 2 \end{bmatrix}$. En $(0, 0)$ el hessiano es $\begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix}$.

FIGURA 9.1.2. Gráfica de la función $F(x, z) = 2x^2z^2 - 4z - 4x$ FIGURA 9.1.3. Gráfica de $F(x, y) = xe^x \cos y$

Para calcular los valores propios se resuelve el polinomio característico $(2 - \lambda)^2 - 4$. En este caso las raíces son 0 y 4, es decir, la matriz es semidefinida positiva y no se puede determinar si $(0, 0)$ es mínimo, máximo o punto de silla. Se sabe que no puede ser máximo porque la función crece en la dirección asociada al valor propio 4, que en este caso corresponde al vector propio $(-1, 1)$ ¹, pero no se puede indicar que se trata de un mínimo porque en este caso el término infinitesimal

¹Nótese que la dirección donde crece la función es con respecto al punto estacionario. Por ejemplo, si el punto estacionario hubiese sido $(3, 2)$ en lugar de $(0, 0)$ la dirección sería en función de $(2, 2)$ y no de $(0, 0)$, es decir, sería la dirección de la recta que pasa por los puntos $(4, 3)$ y $(3, 2)$.

FIGURA 9.1.4. Gráfica de $F(x, y) = (x - y)^2 + x^3$

que normalmente se desprecia en la aproximación dada por el teorema de Taylor ya no se puede despreciar.

La Figura 9.1.4 muestra que el punto $(0, 0)$ es realmente un punto de silla, pero eso no se puede determinar a partir del vector propio asociado al valor propio 0.

Sustituyendo el punto $x_0 = (0, 0)$ y los valores de la función, gradiente y hessiano en ecuación 9.1.1, la aproximación de Taylor para F es

$$F(x) = \frac{1}{2} \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \epsilon$$

$$F(x) = x^2 - 2xy + y^2 + \epsilon = (x - y)^2 + \epsilon$$

Como puede verse, el efecto del término de tercer orden es despreciado en la aproximación dada por el teorema de Taylor y su influencia se desconoce en el caso del valor propio 0.

En este caso, se puede determinar el valor del término ϵ mediante el análisis de la diferencia entre $F(x)$ y la aproximación cuadrática. En este caso, $\epsilon = (x - y)^2 + x^3 - (x - y)^2 = x^3$, la cual representa un punto de inflexión en el punto $(0, 0)$. Recordar que en una función de una variable, suponiendo que la primera derivada distinta de cero es de orden n , si n es par se trata de un extremo (si el signo es positivo indica un mínimo y negativo indica máximo) y si n es impar se trata de un punto de inflexión. En este caso la primera derivada distinta de cero es la tercera derivada, ya que $e'(x) = 3x^2 = 0$, $e''(x) = 6x = 0$ y $e'''(x) = 6$ y por lo tanto en $(0, 0)$ $F(x)$ tiene un punto de silla.

DESCOMPOSICIÓN EN VALORES SINGULARES (SVD)

10.1. Introducción

Cualquier matriz A $m \times n$ puede ser factorizada

$$(10.1.1) \quad A = U \Sigma V^T$$

donde U, V son ortogonales y Σ $m \times n$ sólo tiene elementos positivos o nulos en su diagonal.

Los elementos σ_i de la diagonal de Σ están ordenados de forma decreciente y los últimos $n - r$ términos son cero. Estos valores reciben el nombre de valores singulares y cumplen

$$(10.1.2) \quad A v_i = \sigma_i u_i$$

Donde los v_i y los u_i son las columnas de V y U respectivamente.

La matriz U es una base ortonormal para \mathbb{R}^m , la matriz V es una base ortonormal para \mathbb{R}^n y geoméricamente puede interpretarse la matriz A como una transformación de un vector x del espacio \mathbb{R}^m a \mathbb{R}^n en la cual cada componente de x en la base V tiene asociada una dilatación en cada componente de la base U . El valor singular σ_i indica el factor de la dilatación.

De 10.1.1,

$$A = [u_1 u_2 \cdots u_m] \begin{bmatrix} \sigma_1 & & & & & \\ & \sigma_1 & & & & \\ & & \ddots & & & \\ & & & \sigma_r & & \\ & & & & \ddots & \\ & & & & & \sigma_n \\ & & & & & & \vdots \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix}$$

entonces

$$A = \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \cdots & \sigma_r u_r & \cdots & 0 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix} = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

y

$$(10.1.3) \quad A = U \Sigma V^T = \sum_{i=1}^r \sigma_i u_i v_i^T$$

Nótese que los σ_i están ordenados decrecientemente y que cada u_i y v_i es normalizado, entonces a medida que i aumenta, la influencia de $\sigma_i u_i v_i^T$ sobre el valor de A disminuye.

10.2. Aplicación a compresión de datos

El resultado anterior puede utilizarse en aplicaciones de compresión de datos en las cuales se puede aproximar el valor de A utilizando los primeros h elementos de la sumatoria 10.1.3. La ventaja de este enfoque radica en el hecho de que una matriz que requiere mn números puede ser representada de manera aproximada utilizando $h + mh + nh = h(1 + m + n)$ números, lo cual puede representar una reducción significativa en el tamaño de los datos.

Por ejemplo, una imagen en escala de grises puede ser representada mediante una matriz donde cada elemento representa el color (por ejemplo, 0 negro y 255 blanco)¹ de cada pixel². Por ejemplo, una imagen de 1000×1000 pixels requiere un millón de valores para ser representada, pero utilizando los primeros 100 valores singulares podría obtenerse una muy buena aproximación y se requerirían únicamente $100(1 + 1000 + 1000) = 200100$ valores lo cual representaría una compresión de casi el 80 %.

10.3. Construcción de la SVD

Supóngase que se tiene una matriz rectangular A de $m \times n$. Entonces, la matriz simétrica $A^T A$ de $n \times n$ tiene un conjunto de vectores propios ortonormales x_1, x_2, \dots, x_n . Este conjunto de vectores es una base de \mathbb{R}^n y constituirán las columnas de la matriz V .

$$(10.3.1) \quad A^T A x_j = \lambda_j x_j \quad \text{con } x_j^T x_j = 1 \text{ y } x_i^T x_j = 0 \text{ para } i \neq j$$

Si se calcula el producto escalar con x_j se obtiene

$$(10.3.2) \quad x_j^T A^T A x_j = \lambda_j x_j^T x_j$$

de donde

$$(10.3.3) \quad \|A x_j\|^2 = \lambda_j$$

por lo tanto $\lambda_j \geq 0$.

Supóngase que se decide organizar los x_j y λ_j de manera tal que los λ_j estén ordenados de manera decreciente. En este caso se puede asumir que los primeros r valores propios son positivos y que los últimos $n - r$ son cero.

Para los $\lambda_j > 0$ sean

$$(10.3.4) \quad \sigma_j = \sqrt{\lambda_j} \text{ y } q_j = A x_j / \sigma_j$$

Estos q_j son vectores unitarios (según ecuación 10.3.3, σ_j es la norma de $A x_j$) y ortogonales en \mathbb{R}^m . Según las ecuaciones 10.3.1 y 10.3.2:

$$(10.3.5) \quad q_i^T q_j = \frac{x_i^T A^T A x_j}{\sigma_i \sigma_j} = \frac{\lambda_j x_i^T x_j}{\sigma_i \sigma_j} = 0$$

Estos r vectores ortonormales q_j pueden ser extendidos (de ser necesario) a m vectores ortonormales mediante el procedimiento de Gram-Schmidt, formando así una base de \mathbb{R}^m . Estos vectores q_1, q_2, \dots, q_m serán las columnas de la matriz V .

Considérese ahora el producto $\Sigma = U^T A V$. El elemento en la posición i, j de este producto es $q_i^T A x_j$.

¹una imagen a color suele ser representada mediante tres matrices conteniendo cada una la intensidad de los colores rojo, verde y azul.

²Pixel: picture element o elemento de imagen.

$$(10.3.6) \quad q_i^T A x_j = \begin{cases} 0 & \text{si } j > r \text{ porque } A x_j = 0 \\ q_i^T \sigma_j q_j & \text{si } j \leq r \text{ porque } A x_j = \sigma_j q_j \end{cases}$$

Pero $q_i^T q_j = 0$ excepto en el caso $i = j$, en cuyo caso es 1. Entonces, las únicas entradas distintas de cero en la matriz $\Sigma = U^T A V$ son los primeros r elementos de la diagonal, los cuales son $\sigma_1, \sigma_2, \dots, \sigma_r$. Como U y V son ortogonales entonces

$$A = U \Sigma V^T$$